

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: November 22, 2002, 13:29:17 ; Search time 16 seconds
(without alignments)
891.882 Million cell updates/sec

Title: US-09-854-562-2

Perfect score: 2519
Sequence: 1 MKSFNTEGHNHSTAEGDAY.....SIAGLISSVKYKPPRTNHE 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCtUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2519	100.0	485	1	US-08-362-512A-2
2	2519	100.0	485	4	US-08-964-939-2
3	1501	59.6	493	1	US-08-362-512A-4
4	1501	59.6	493	4	US-08-964-939-4
5	567	22.5	447	4	US-09-370-253-6
6	563.5	22.4	447	4	US-09-370-253-10
7	527	20.9	447	4	US-09-370-253-2
8	379	15.0	259	4	US-09-370-253-4
9	310	12.3	432	4	US-09-370-253-12
10	310	12.3	473	4	US-09-637-118B-2
11	198.5	7.9	451	1	US-08-191-337-3
12	121.5	4.8	125	4	US-09-370-253-16
13	110	4.4	494	4	US-09-134-001C-4475
14	108	4.3	1212	4	US-09-268-866-2
15	104.5	4.1	438	2	US-08-677-049-9
16	104	4.1	256	4	US-09-351-224E-8
17	102.5	4.1	525	4	US-09-134-001C-4648
18	101.5	4.0	482	4	US-09-134-001C-4309
19	97.5	3.9	2516	3	US-08-374-077C-2
20	97.5	3.9	2516	4	US-08-895-590-2
21	97.5	3.9	2516	4	US-09-539-879A-2
22	97	3.9	467	4	US-09-134-001C-3020
23	97	3.9	489	4	US-09-134-001C-5241
24	97	3.9	521	4	US-09-134-001C-4290
25	96	3.8	518	4	US-09-134-001C-4744
26	95.5	3.8	326	4	US-08-986-768-2
27	95.5	3.8	797	4	US-09-191-468-120

28	95.5	3.8	797	4	US-09-191-468-122	Sequence 122, App
29	93.5	3.7	347	1	US-08-118-270-47	Sequence 47, Appl
30	93.5	3.7	347	5	PCR-US93-08528-47	Sequence 47, Appl
31	93.5	3.7	2522	4	US-09-251-645-13	Sequence 13, Appl
32	93	3.7	424	4	US-09-134-001C-5009	Sequence 5009, App
33	92.5	3.7	540	4	US-09-433-994-2	Sequence 2, Appl1
34	92.5	3.7	622	2	US-08-132-990A-4	Sequence 4, Appl1
35	92.5	3.7	622	5	PCR-US92-09382-4	Sequence 4, Appl1
36	92	3.7	109	4	US-09-370-253-8	Sequence 8, Appl1
37	92	3.7	496	4	US-09-134-001C-3001	Sequence 3001, App
38	92	3.7	511	2	US-09-073-362-1	Sequence 1, Appl1
39	92	3.7	511	2	US-09-243-920-1	Sequence 1, Appl1
40	90	3.6	349	4	US-09-134-001C-4004	Sequence 4004, Ap
41	90	3.6	797	2	US-08-700-013B-19	Sequence 19, Appl
42	90	3.6	797	2	US-08-700-013B-21	Sequence 21, Appl
43	90	3.6	797	4	US-09-191-468-124	Sequence 124, App
44	89.5	3.6	368	4	US-08-709-838-2	Sequence 2, Appl1
45	89.5	3.6	368	4	US-08-829-839-2	Sequence 2, Appl1

ALIGNMENTS

RESULT 1
US-08-362-512A-2
; Sequence 2, Application US/08362512A
; Patent No. 5719043
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Oscrotenk, Faber, Gerd & Seffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/362,512A
; FILING DATE: 05-JAN-1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Weisman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700
; TELEFAX: (212) 382-0888
; TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-362-512A-2
Query Match 100.0%; Score 2519; DB 1; Length 485;
Best Local Similarity 100.0%; Pred.No. 1.2e-261;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MKSFNTEGHNHSTAEGDAYTSDPTKXNVEDGDERKRTGTWLTASAHITAVIGSGVLSL 60

1 M K S F N T E G N H S T A E S G D A Y T V S D P T K N V D E D G R E K R T G T W L T A S A H I I T A V I G S G V L S L 60
61 A W A I A Q L G W I A G T S I L L I F S F T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G G R K 120
61 A W A I A Q L G W I A G T S I L L I F S F T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G G R K 120
121 V Q L C G V A Q Y G N L I G V T V G T T I T A S I S L V A V G K S N C F H D K G H T A D C T I S N Y P Y M A V F G I I Q 180
121 V Q L C G V A Q Y G N L I G V T V G T T I T A S I S L V A V G K S N C F H D K G H T A D C T I S N Y P Y M A V F G I I Q 180
181 V I L S Q I P N P H K L S F L S I M A A V M S F T Y A T I G I G L A T A T V A G G K V G K T S M T G T A V G D V T A A 240
181 V I L S Q I P N P H K L S F L S I M A A V M S F T Y A T I G I G L A T A T V A G G K V G K T S M T G T A V G D V T A A 240
241 Q K I W R S F O A V G D I A F A Y A V A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300
241 Q K I W R S F O A V G D I A F A Y A V A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300
301 I G Y A A F G N N A P G D F L T D F G F F E P F W L I D F A N A C I A V H L I G A Y O V F A Q P I F Q F V E K K C N R N 360
301 I G Y A A F G N N A P G D F L T D F G F F E P F W L I D F A N A C I A V H L I G A Y O V F A Q P I F Q F V E K K C N R N 360
361 Y P D N K F I T S E Y S V N V P F L G K N I S F R L V R T A Y V V I T T V A M I F P F F N A I L G L I G A A S F 420
361 Y P D N K F I T S E Y S V N V P F L G K N I S F R L V R T A Y V V I T T V A M I F P F F N A I L G L I G A A S F 420
421 W P L T V Y F P V E M H I A Q T K I K K Y S A R M I A L K T M C Y V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480
421 W P L T V Y F P V E M H I A Q T K I K K Y S A R M I A L K T M C Y V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480
481 R T M H E 485
481 R T M H E 485

RESULT 2
US-08-964-939-2
; Sequence 2, Application US/08964939
; Patent No. 6245970
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/964,939
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/362,512
; FILING DATE: 05-JAN-1995
; APPLICATION NUMBER: US 08/007,636
; FILING DATE: 21-JAN-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Meilman, Edward A.
; REGISTRATION NUMBER: 24,735
; REFERENCE/DOCKET NUMBER: P/951-107
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888
TELEX: 236925
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 485 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-964-939-2
Query Match 100.0%; Score 2519; DB 4; Length 485;
Best Local Similarity 100.0%; Pred. No. 1.2e-261;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 M K S F N T E G N H S T A E S G D A Y T V S D P T K N V D E D G R E K R T G T W L T A S A H I I T A V I G S G V L S L 60
Db 1 M K S F N T E G N H S T A E S G D A Y T V S D P T K N V D E D G R E K R T G T W L T A S A H I I T A V I G S G V L S L 60
QY 61 A W A I A Q L G W I A G T S I L L I F S F T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G G R K 120
Db 61 A W A I A Q L G W I A G T S I L L I F S F T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G G R K 120
QY 121 V Q L C G V A Q Y G N L I G V T V G T T I T A S I S L V A V G K S N C F H D K G H T A D C T I S N Y P Y M A V F G I I Q 180
Db 121 V Q L C G V A Q Y G N L I G V T V G T T I T A S I S L V A V G K S N C F H D K G H T A D C T I S N Y P Y M A V F G I I Q 180
QY 181 V I L S Q I P N P H K L S F L S I M A A V M S F T Y A T I G I G L A T A T V A G G K V G K T S M T G T A V G D V T A A 240
Db 181 V I L S Q I P N P H K L S F L S I M A A V M S F T Y A T I G I G L A T A T V A G G K V G K T S M T G T A V G D V T A A 240
QY 241 Q K I W R S F O A V G D I A F A Y A V A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300
Db 241 Q K I W R S F O A V G D I A F A Y A V A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300
QY 301 I G Y A A F G N N A P G D F L T D F G F F E P F W L I D F A N A C I A V H L I G A Y O V F A Q P I F Q F V E K K C N R N 360
Db 301 I G Y A A F G N N A P G D F L T D F G F F E P F W L I D F A N A C I A V H L I G A Y O V F A Q P I F Q F V E K K C N R N 360
QY 361 Y P D N K F I T S E Y S V N V P F L G K N I S F R L V R T A Y V V I T T V A M I F P F F N A I L G L I G A A S F 420
Db 361 Y P D N K F I T S E Y S V N V P F L G K N I S F R L V R T A Y V V I T T V A M I F P F F N A I L G L I G A A S F 420
QY 421 W P L T V Y F P V E M H I A Q T K I K K Y S A R M I A L K T M C Y V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480
Db 421 W P L T V Y F P V E M H I A Q T K I K K Y S A R M I A L K T M C Y V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480
QY 481 R T M H E 485
Db 481 R T M H E 485

RESULT 3
US-08-362-512A-4
; Sequence 4, Application US/08362512A
; Patent No. 5719043
; GENERAL INFORMATION:
; APPLICANT: FROMMER, Wolf-Bernd
; TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
; TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
; TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
; STREET: 1180 Avenue of the Americas
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10036-8403
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/362,512A
FILING DATE: 05-JAN-1995
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/007,636
FILING DATE: 21-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-107
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-362-512A-4

Query Match 59.6%; Score 1501; DB 1; Length 493;
Best Local Similarity 56.7%; Pred. No. 2.1e-152;
Matches 271; Conservative 94; Mismatches 111; Indels 2; Gaps 2;

QY 8 GHNHSTAESGDVATVSDPKNVDEGRKRTGTWLTASAHITTAIVGSGVLSAMIAQL 67
DB 17 GHQVFDVASHDEVPPQPAKCFDDGRLKRTGTWLTASAHITTAIVGSGVLSAMIAQL 76
QY 68 GWIAGTSILILSFITYFTSTMLADCYRAPDPYTGKNTYMDVNSYLGGRVQLCGVA 127
DB 77 GWIAGRAWMLLFSVLYLSTSLSDCYRGDANSKGRNTYMDAVSILGGRFKICGLI 136
QY 128 QYGNLIGTVGYTTASISLVAVGKSCNCFHDKGHTADCTISNYPYNAVFGIIVILSQIP 187
DB 137 QYLNFGIAGIYTTASISIMMAIKRSCNCFHSGKDPCHMSSNPYMIIVFGVAEILLSQVP 196
QY 188 NPHKLSFLSMAVMSFTYATIGIATVAGKTKGKTSMTGTAVGVDTAOKITWRSF 247
DB 197 DFDQIWMISIVAAVMSFTYSAIGLGIQVAVANGVFKSLGTISIGT-VTQTKIWRTP 255
QY 248 QAVGDIAFAYAVATVLEIYODTLRSSPAENKAMKRASLVGVSFTTFYILCGCIGYAAFG 307
DB 256 QALGDIAPFAYSISVULIEIODTVRSPPASKTKKATKISIAVTTTFYMLCGSMGTAAFG 315
QY 308 NNAPGDLTDGFFEFEPWLIIDFANACIAVHLIGAYOVFAQPIFOVEKKCNRYPNPKFI 367
DB 316 DAAPGNLTLGFGFYNPFLDIDNNAIIVHLVGAYOVFAQPIFAPIEKSVAERYPDNDFL 375
QY 368 TSEYSVNP-FLGKNISLFLVWRTAVYVITTVAMIRPFPAIILGLIGAASFPLTYV 426
DB 376 SKFEIRIQFQSPYKVNFRMYRSGFVYTTVVISMLMFPFNDVVGIIIGALFHWPLTYV 435
QY 427 FPEVEMIAOTKIKKYSARWIALKTMVCYCLIVSLAAGSIAGLISVKTYPKPTMH 484
DB 436 FPEVEMTIKQKVKMSTRWVCLQMLSVACLVISVAVGVSIGAVMDLKVYKPFKSTY 493

RESULT 4
US-08-964-939-4
Sequence 4, Application US/08964939
Patent No. 6245970
GENERAL INFORMATION:
APPLICANT: FROMMER, Wolf-Bernd
TITLE OF INVENTION: DNA SEQUENCES FOR AN AMINO ACID
TITLE OF INVENTION: TRANSPORTER, PLASMIDS, BACTERIA, YEASTS AND PLANTS
TITLE OF INVENTION: CONTAINING A TRANSPORTER AND THEIR USE
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSES: Ostrolenko, Faber & Soffen
STREET: 1160 Avenue of the Americas

CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/964,939
FILING DATE:
CLASSIFICATION: 800
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,512
FILING DATE: 05-JAN-1995
APPLICATION NUMBER: US 08/007,636
FILING DATE: 21-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/951-107
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 493 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-964-939-4

Query Match 59.6%; Score 1501; DB 4; Length 493;
Best Local Similarity 56.7%; Pred. No. 2.1e-152;
Matches 271; Conservative 94; Mismatches 111; Indels 2; Gaps 2;

QY 8 GHNHSTAESGDVATVSDPKNVDEGRKRTGTWLTASAHITTAIVGSGVLSAMIAQL 67
DB 17 GHQVFDVASHDEVPPQPAKCFDDGRLKRTGTWLTASAHITTAIVGSGVLSAMIAQL 76
QY 68 GWIAGTSILILSFITYFTSTMLADCYRAPDPYTGKNTYMDVNSYLGGRVQLCGVA 127
DB 77 GWIAGRAWMLLFSVLYLSTSLSDCYRGDANSKGRNTYMDAVSILGGRFKICGLI 136
QY 128 QYGNLIGTVGYTTASISLVAVGKSCNCFHDKGHTADCTISNYPYNAVFGIIVILSQIP 187
DB 137 QYLNFGIAGIYTTASISIMMAIKRSCNCFHSGKDPCHMSSNPYMIIVFGVAEILLSQVP 196
QY 188 NPHKLSFLSMAVMSFTYATIGIATVAGKTKGKTSMTGTAVGVDTAOKITWRSF 247
DB 197 DFDQIWMISIVAAVMSFTYSAIGLGIQVAVANGVFKSLGTISIGT-VTQTKIWRTP 255
QY 248 QAVGDIAFAYAVATVLEIYODTLRSSPAENKAMKRASLVGVSFTTFYILCGCIGYAAFG 307
DB 256 QALGDIAPFAYSISVULIEIODTVRSPPASKTKKATKISIAVTTTFYMLCGSMGTAAFG 315
QY 308 NNAPGDLTDGFFEFEPWLIIDFANACIAVHLIGAYOVFAQPIFOVEKKCNRYPNPKFI 367
DB 316 DAAPGNLTLGFGFYNPFLDIDNNAIIVHLVGAYOVFAQPIFAPIEKSVAERYPDNDFL 375
QY 368 TSEYSVNP-FLGKNISLFLVWRTAVYVITTVAMIRPFPAIILGLIGAASFPLTYV 426
DB 376 SKFEIRIQFQSPYKVNFRMYRSGFVYTTVVISMLMFPFNDVVGIIIGALFHWPLTYV 435
QY 427 FPEVEMIAOTKIKKYSARWIALKTMVCYCLIVSLAAGSIAGLISVKTYPKPTMH 484
DB 436 FPEVEMTIKQKVKMSTRWVCLQMLSVACLVISVAVGVSIGAVMDLKVYKPFKSTY 493

RESULT 5
US-09-370-253-6

; Sequence 6, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
; CURRENT APPLICATION NUMBER: US/09/370,253
; EARLIER FILING DATE: 1999-08-09
; EARLIER FILING DATE: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 6
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-370-253-6

Query Match 22.5%; Score 567; DB 4; Length 447;
Best Local Similarity 31.9%; Pred. No. 3.2e-52;
Matches 153; Conservative 81; Mismatches 167; Indels 78; Gaps 18;

QY 31 EDGR-----EKRTGTLTASAHITAVIGSGVLSLAWAIAQLGWIAGTSL 76
DB 13 KGRSAQAEKADLDLPITSSRNAKWMYSAFHNVTAMVGAGVLSLPYAMSELGWPGLAVL 72
QY 77 LIFSFTYFTSTMLADCYRAPDPVTGKRNVTYMDVVRSLGGRKVLQGVV--QYGNLIG 134
DB 73 ILSWIITLTLQWVEMH---EMVPGKRFDRYHE--LGQHAFGKGLGLWVVPQQLVVEVG 128
QY 135 VTVGYTITASISLVAVGKSNCFHD---KGTADCTISNYPYMAVFIQVILSQIPNFK 191
DB 129 VNIIVYMTGKSL-----KPFHDVLCGCKNKLTY--FIMIFASVHFVLSQLPNLN 181
QY 192 LQFLSTMAAVMSFTYATIGLAIAIVAGKVGKTSMTGTAVGVDTAAQKWRSPQAVG 251
DB 182 ISGVSLAAAVMSLSYSTIANG---ASVDKGVADVHYHRA-----TTSKVGKFGPSALG 234
QY 252 DIAFAVAYATVLEIOTLRSSPAENKAMKRASLVGVSTTTFFYILC----CGIGYAAGF 307
DB 235 DVAFAYAGHNVLLEIQTIPSTP--EKPSKPMKMGVWVAVIIVVVALCYPPVALGVYAFG 292
QY 308 NNAPGDFLTDFGFEFFWLDIFANACIAVHLIGAYQVFAQPIQFVE----KKCNRYPD 363
DB 293 NHVDDNLTLS--RPKWLIALANMMVVIHVIGSYQIYAMPVDFMIEIETVLVKKL--RFP- 347
QY 364 NKPFITSESVNVPFLGKFNISLFRLVWRTAYVITTVVAMIPFFNAIILGLIGAASF 423
DB 348 -----PGL-----TLRLIARTLYVAFTMFIATITPFFGGLGFFGGFAFAPT 389
QY 424 TVVFPVEMHIAOTKIKKYSGARMIAKTMCVCLIVS-----LLAAAGSIAGLISSVKTYK 478
DB 390 TYFLPCIMWLAIYKPRFSLSWFT-----NWCIILGVLMLSPIGGLRQIILDAIKTYK 444

RESULT 6
US-09-370-253-10
; Sequence 10, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
; CURRENT APPLICATION NUMBER: US/09/370,253
; EARLIER FILING DATE: 1999-08-09
; EARLIER FILING DATE: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97

; SEQ ID NO 10
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Triticum aestivum
US-09-370-253-10

Query Match 22.4%; Score 563.5; DB 4; Length 447;
Best Local Similarity 32.1%; Pred. No. 7.5e-52;
Matches 154; Conservative 85; Mismatches 168; Indels 73; Gaps 16;

QY 25 PTKNVDEGREK-----RGTWLTASAHITAVIGSGVLSLAWAIAQLGWIAGTSL 74
DB 12 PAKDVRTDEQKKIDDLPITSSRNAKWMYSAFHNVTAMVGAGVLSLPYAMSELGWPGLIA 71
QY 75 ILLIFSFTYFTSTMLADCYRAPDPVTGKRNVTYMDVVRSLGGRKVLQGVVAVQYGNL-I 133
DB 72 VLVISWVITLTLQWVEMH---EMVPGKRFDRYHELGQHAFGDKGLWVVPQQLIIEV 128
QY 134 GVTGYTITASISLVAVGKSNCFHD---KGTADCTISNYPYMAVFIQVILSQIPNFK 190
DB 129 GVNIVYMTGGRSL-----KPFHDVICGKCKDKLTF--FIMIFASVHFVLSQLPNLN 180
QY 191 KLSFLUSMAAVMSFTYATIGLAIAIVAGKVGKTSMTGTAVGVV----VTAAQKIWR 245
DB 181 SISGVSLAAAVMSLSYSTIANG---ASVDKGM-----VDVYNLRATTTPEKVFEG 228
QY 246 SFQAVGDIATAYATVLEIOTLRSSPAE--NKAMKRASLVGVSTTTFFYILCGCIGY 303
DB 229 FFGALGEVAFAYAGHNVLLEIQTIPSTPEKPSKPMKMGVWVAVIIVVVALCYPPVALIGY 288
QY 304 AAFGNNAQDFLTDFGFEFFWLDIFANACIAVHLIGAYQVFAQPIQFVEKKCNRYPD 363
DB 289 WAFGNSVDNLTILN--KPTWLISTANMMVVIHVIGSYQIYAMPVDFMIE-----337
QY 364 NKPFITSESVNVPFLGKFNISLFRLVWRTAYVITTVVAMIPFFNAIILGLIGAASF 422
DB 338 -----TVLVKKL-KFPFGLTLRLIARTLYVAFTMFIATITPFFGGLGFFGGFAFAP 388
QY 423 LTVFPVEMHIAOTKIKKYSGARMIAKTMCVCLIVS-----LLAAAGSIAGLISSVKTYK 478
DB 389 TYFLPCIMWLAIYKPRFSLSWFT-----NWCIILGVLMLSPIGGLRQIILDSKTYK 444

RESULT 7
US-09-370-253-2
; Sequence 2, Application US/09370253
; Patent No. 6165792
; GENERAL INFORMATION:
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
; CURRENT APPLICATION NUMBER: US/09/370,253
; CURRENT FILING DATE: 1999-08-09
; EARLIER FILING DATE: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 2
; LENGTH: 447
; TYPE: PRT
; ORGANISM: Hordeum vulgare
US-09-370-253-2

Query Match 20.9%; Score 527; DB 4; Length 447;
Best Local Similarity 28.9%; Pred. No. 6.2e-48;
Matches 142; Conservative 89; Mismatches 177; Indels 84; Gaps 18;

QY 21 TVSDPTKNV----DEGREKRTGTWLTASA-----HIITAVIGSGVLSLAWAIA 65
DB 3 THASPDNTPPKDERTAREKAIDDLPTITSSRNAKWMYSAFHNVTAMVGAGVLSLPYAMS 62


```
; APPLICANT: Allen, Stephen M.
; APPLICANT: Sakai, Hajime
; APPLICANT: Thorpe, Catherine J.
; TITLE OF INVENTION: Amino Acid Transporters
; FILE REFERENCE: BB-1200
; CURRENT APPLICATION NUMBER: US/09/370,253
; CURRENT FILING DATE: 1999-08-09
; EARLIER APPLICATION NUMBER: 60/097,222
; EARLIER FILING DATE: August 20, 1998
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 16
; LENGTH: 125
; TYPE: PRT
; ORGANISM: Glycine max
US-09-370-253-16

Query Match      4.8%; Score 121.5; DB 4; Length 125;
Best Local Similarity 27.3%; Pred. No. 2.9e-05;
Matches 39; Conservative 27; Mismatches 46; Indels 31; Gaps 7;

Qy 346 AADPQFVKCKNBNYNDKFTSESVNVPFLGKFNISLFLWMTAVVITTVAMIF 405
Db 3 ASPMTEYLDTK-----YIGKSAALAFNLS-FRVLVGGYLVTFVVSALL 47
Qy 406 PEFNALIGLGAASFPLTVFPEVEMHIAQTIKIKYSAR---WIAKTMCVCLIVSL 461
Db 48 PFLGDMSLTGAISTPFLFILANHYLVITNENKLTSTOKLMMI---NICF---FALM 100
Qy 462 AAAGSIAGL---ISSVKTYPK 480
Db 101 SAAAMIAALRLIDLS-KTYHVF 122

RESULT 13
US-09-134-001C-4475
; Sequence 4475, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucet-Re-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4475
; LENGTH: 494
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-4475

Query Match      4.4%; Score 110; DB 4; Length 494;
Best Local Similarity 21.5%; Pred. No. 0.0041;
Matches 90; Conservative 52; Mismatches 124; Indels 152; Gaps 23;

Qy 164 DCTISNYPMAVFGIIOVLISQIPNFHKLSTLSMAVMSFT--VATTIGLAIVTA-- 219
Db 67 DGSISYGLIAVITIAI-----FSLNTYMSV-KLITSFTLSISVFKGIVILITIML 118
Qy 220 -----GKVKTSMTGTAVGVDTVAQAQIKMSFOAVGDIAPAVAYATVL---IEI 266
Db 119 LVSGFDTGVGHSIGTFMPYGSAPIFAATTSGIISFN-----AFQTIINMSSEI 169
Qy 267 QDTLRSSPAENKMKRASLVGVSTTTFYLLCGCIGAAFGNNAAPGFLIDFG-----FP 321
Db 170 K---NPEKNIA--RGVISTLSALIVTL---OSTYITSPSSMLHEHSGINFN 218
Qy 322 EPF-----WLIDFANACIAVHLIGAYQVFAQPI--PQVFEKKCNRYPNKFTISE 370
```

```
Db 219 SPFADMAILGNWL-----AIILYMEAVSPFGTGSFVAVTGRVLRAMEENGHIK 271
Qy 371 YSVNVPELGFKN-----ISLFLWMTAVVYI-----T 398
Db 272 -----FLGKIKKNINIPVATAFNALISMVWTLFR-DWGLAVISTATLVAYITGTF 324
Qy 399 TVVA-----MIPFENALIGLGAASF-----WPLT-----VYFP 428
Db 325 TVISLRKAPKQTRPPKANIILKFNAPLSFVLASLAIVYMMWFTTAEVILIIILGLPIYFF 384
Qy 429 VE--WHIAQTIKIKYSARIALKTMCVCLIVSLAAAGSINGLSSVTTYPRFTMH 484
Db 385 YEKQKWKQNTKKQIGGSLWII---YLIVLAFISFGS-----KEFGLNWIH 429

RESULT 14
US-09-268-866-2
; Sequence 2, Application US/09268866
; Patent No. 6316272
; GENERAL INFORMATION:
; APPLICANT: Mack, David
; APPLICANT: Gish, Kurt C.
; APPLICANT: Wilson, Keith E.
; TITLE OF INVENTION: No. 6316272el Methods of Diagnosing of Colorectal Cancer,
; TITLE OF INVENTION: Compositions, and Methods of Screening for Colorectal
; FILE REFERENCE: A67474/BMS/DAV/JJD
; CURRENT APPLICATION NUMBER: US/09/268,866
; CURRENT FILING DATE: 1999-03-15
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-866-2

Query Match      4.3%; Score 108; DB 4; Length 1212;
Best Local Similarity 17.7%; Pred. No. 0.027;
Matches 110; Conservative 87; Mismatches 232; Indels 194; Gaps 27;

Qy 6 TEGHN-----HSTAESGD-----AYTVSDPKVNDDEGRKRTGTWLTASA 46
Db 212 TGHNTMDAVPRIDHYRHTAQLGEXLNPSLAEHDELEKEPFEDGFANGSESTRTDA 271
Qy 47 HITTAIVSGGVLSTLW-----AIAOLGMA-----GTSILIFSTITYF 85
Db 272 VVYTAESKGVVKGFGIKGVLRCLMNLINGWMLFRLSWIVGAGIGLSVLVIMATVYT 311
Qy 86 TSTMLADCYRAPDP-VTGRKNTYMDVVRVSYLGRVQJGVAVQGNLGVTVGYTITAS 144
Db 332 TTIGLSTSAIATNGFVRGGAY-----YLISRLG---PEFGAIGLIFAFANAVA 379
Qy 145 ISLVAGKSNCFHD-KGHTADCTISNYPMAVFGIIOVLISQIPNFHKLSTLSMAVVA 202
Db 380 VMYVVGFAETVELLEKESI-LMIDEINDIRIGAITVI-----LLGISVAGM 428
Qy 203 SEFYATITIGLAIVTAGK--VG-----KTSMTGTAVGVDTVAQAQIKW 244
Db 429 EWEAKQIYLVILVILALIDFVIGTFIPLESKKPGFPGYKSEIFENENGPDREREETFP 488
Qy 245 RSFOAGDIAFAVAVTVLIEIOTDIRSSPAENKMKRASLVGVSTTTFYI-----LCG 299
Db 489 SYF-----AIFPAATGILAGANISGLADPOSALPKGTLLAILITLVYGVIAVSGS 542
Qy 300 CIGYAAFGN-----NAPG-----DFGFPFPFLWIDFAN----- 331
Db 543 CVVRDATGVNDIYVELNCTSAACKLNFDFSSCSSPSCSYGLMNNFQVMSVSGFTPL 602
Qy 332 -----ACIAVHLIGAYQV-----AQPIQVFEKKCNRYPNK----- 365
Db 603 ISAGIFSATLSALASLVAPKIFQALCKDNTYPAFQWPAKGVGKNNBPLRGVILTFLIA 662
```

QY 366 ---FITSEYVNVVFLGKFNISLFRIV-----WRTAVVVTITVVAMIEPFF 408
Db 663 LGPILIAELNVIAPITISNFFLASALINVSFVHSLAKSPGRPA-----FKYY 711
QY 409 NAILGIGIA-----ASFWD--LTVVFPVEMHIAQTIKKYSARM-IALKTMICYVCL 456
Db 712 NMWISILGAILCCIVMFVNWAAALTYVIVLGLIYVT-YKKPDVNWGSSQALTY--- 767
QY 457 IVSLLAAAGSIAGLISSVTKYP 479
Db 768 -LNALQHSIRLSGVEDHVKNFRP 789

RESULT 15

US-08-677-049-9
; Sequence 9, Application US/08677049
; Patent No. 5858707
; GENERAL INFORMATION:
; APPLICANT: Guimaraes, M. Jorge
; APPLICANT: Bazan, J. Fernando
; APPLICANT: McClanahan, Terrill K.
; APPLICANT: Zlotnik, Albert
; TITLE OF INVENTION: PURIFIED MAMMALIAN NUCLEOBASE PERMEASES;
; TITLE OF INVENTION: NUCLEIC ACIDS, ANTIBODIES
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: DNAX Research Institute
; STREET: 901 California Avenue
; CITY: Palo Alto
; STATE: California
; COUNTRY: USA
; ZIP: 94304-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/677,049
; FILING DATE: 03-JUL-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/000,788
; FILING DATE: 03-JUL-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Ching, Edwin P.
; REGISTRATION NUMBER: 34,090
; REFERENCE/DOCKET NUMBER: DX0511
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415-852-9196
; TELEFAX: 415-496-1200
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 438 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Region
; LOCATION: 115..144
; OTHER INFORMATION: /note= "Encompasses TM 4 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 323..357
; OTHER INFORMATION: /note= "Encompasses TM 9 of Figure
; OTHER INFORMATION: 4"
; FEATURE:
; NAME/KEY: Region
; LOCATION: 359..386
; OTHER INFORMATION: /note= "Encompasses TM 10 of Figure

OTHER INFORMATION: 4"

US-08-677-049-9
Query Match 4.1%; Score 104.5; DB 2; Length 438;
Best Local Similarity 20.5%; Pred. No. 0.013;
Matches 75; Conservative 51; Mismatches 131; Indels 109; Gaps 14;
QY 114 SVLGGRKVLQCGVAQ-----YGNLIGVTGVTITASISLVAVGKSNCFHDKGHTAD 164
Db 42 TYLVSIDIFMCGVATLLQWNSNRRFGIGLPVVLGCTFTAVSPMIAIG----- 88
QY 165 CTISNYPYMAVFG--IIQVILSQIPNFHKLSTLSMAAVMSFTVATIGIGLAIAIVAGGK 222
Db 89 ---SEYGVSTVYGSIIASGILVILISFFFGKLVSEFPVPVTVGSVVTI-IGITLMEVA--- 141
QY 223 VGKTSMTGTAVGVDTAAQKIWRSPQAVGDIATVAYATVILIEIODTLRSSPAENKAMKR 282
Db 142 --MNNMAGGEGSAD-----FGDLSNLALAPTVLSIIIVLL---YRFTKGFIKSV-- 184
QY 283 ASLVGVSTTTTFYILCGCIGYAAFGNNAPGDFLTDGFFEP----- 323
Db 185 SILIGILIGTFTIAYPMGKVQFDNVSDAAVQMIOFFYFGAPSFHAAPITWISIVAIVSLV 244
QY 324 -----FWLIDPANACIA-VHLIGAYQ-----VFAQPIQFQVKKCNRNYPDNKFTISEY 371
Db 245 ESTGVYFALGDLTNRLTEIDLKGYRAEGLAVLLGGIF-----NAPPYTAP 291
QY 372 SVNVPFLGKNISLFRLVWRTAYVVVITTVVAMIPFFF-----NAILGLIGAASF 420
Db 292 SONVGLVQLTGIK-----KNAVIVVTGVILMAFGLPKIAAFTTIIPSAVLGGAMVAMF 345
QY 421 WPLTVY 426
Db 346 GMVIAY 351

Search completed: November 22, 2002, 13:31:50
Job time : 17 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2002, 13:28:42 ; Search time 21 Seconds

(without alignments)
2220.247 Million cell updates/sec

Title: US-09-854-562-2

Perfect score: 2519
Sequence: 1 MKSFNTGHNHSTAESGDAY.....SIAGLISSVYKYKPRFMHE 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR_73: *
1: p1r1: *
2: p1r2: *
3: p1r3: *
4: p1r4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2519	100.0	485	2 A48187	amino acid transpo
2	1974.5	78.4	475	2 T00620	probable amino aci
3	1871	74.3	481	2 T50891	amino acid permeas
4	1589.5	63.1	485	2 T07130	probable amino aci
5	1540.5	61.2	486	2 T10100	amino acid transpo
6	1504	59.7	476	2 H96802	probable amino aci
7	1501	59.6	493	2 S52421	amino acid transpo
8	1406.5	55.8	476	2 C96505	probable amino aci
9	1336	52.6	376	2 T07131	probable amino aci
10	1163	46.2	284	2 T09840	amino acid transpo
11	1153.5	45.8	405	2 A57479	amino acid transpo
12	1142	45.3	396	2 B57479	amino acid transpo
13	1109.5	44.0	411	2 C57479	amino acid transpo
14	966	38.3	284	2 T09843	amino acid transpo
15	557	22.1	441	2 C86378	protein P2LJ9.6 [1
16	507.5	20.1	462	2 T15052	amino acid permeas
17	496	19.7	440	2 F86385	probable lysine an
18	493	19.6	450	2 E96738	hypothetical prote
19	326.5	13.0	442	2 T50687	proline transport
20	323.5	12.8	441	2 T50689	proline transport
21	310	12.3	473	2 T50690	proline transport
22	294.5	11.7	439	2 F96517	hypothetical prote
23	294	11.7	439	2 T50688	proline transport
24	292.5	11.6	436	2 D84782	probable proline t
25	287	11.4	442	2 T50692	proline transport
26	281	11.2	439	2 T47713	proline transport
27	241	9.6	473	2 T04965	amino acid transpo
28	208	8.3	713	2 S58251	probable membrane
29	207.5	8.2	470	2 S47892	neutral amino acid

30	202	8.0	436	2 T05653	amino acid transpo
31	200	7.9	516	2 T48238	hypothetical prote
32	189.5	7.5	692	2 S37976	hypothetical prote
33	181.5	7.2	484	2 T34016	hypothetical prote
34	176	7.0	494	2 T16658	hypothetical prote
35	176	7.0	543	2 T48239	hypothetical prote
36	175	6.9	423	2 T49959	hypothetical prote
37	169.5	6.7	485	2 T01244	hypothetical prote
38	167.5	6.6	602	2 S55188	hypothetical prote
39	165.5	6.6	486	2 T42254	amino acid permeas
40	165	6.6	462	2 S42372	hypothetical prote
41	158	6.3	571	2 T06737	hypothetical prote
42	157.5	6.3	529	2 E84813	hypothetical prote
43	152.5	6.1	890	2 T21000	hypothetical prote
44	151.5	6.0	449	2 H88022	protein T27A1.5 [1
45	150.5	6.0	434	2 T15799	hypothetical prote

ALIGNMENTS

RESULT 1

A48187
amino acid transport protein I - Arabidopsis thaliana
N:Alternate names: amino acid permease; neutral amino acid carrier
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 27-Jun-1994 #sequence revision 27-Jun-1994 #ext_change 20-Aug-1999
C:Accession: A48187; A47738; S31970
R:Frommer, W.B.; Hummel, S.; Riemer, J.W.
Proc. Natl. Acad. Sci. U.S.A. 90, 5944-5948, 1993
A:Title: Expression cloning in yeast of a cDNA encoding a broad specificity amino acid
A:Reference number: A48187; PMID:93317600; PMID:8327465.
A:Accession: A48187
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-485 <RNU>
A:Cross-References: EMBL:X67124; NID:G22640; PIDN:CAA32726.1; PID:G22641
A:Note: authors failed to translate the codon ATG for residue 109 as Met
R:Hsu, L.C.; Chlou, T.J.; Chen, L.; Bush, D.R.
Proc. Natl. Acad. Sci. U.S.A. 90, 7441-7445, 1993
A:Title: Cloning a plant amino acid transporter by functional complementation of a yeast
A:Reference number: A47738; PMID:93361464; PMID:8356039
A:Contents: Columbia ecotype
A:Accession: A47738
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-485 <RNU>
A:Cross-References: GB:L16240; NID:G404018; PIDN:AAA32726.1; PID:G404019
A:Note: sequence extracted from NCBI backbone (NCBIN:136517, NCBIIP:136518)
C:Superfamily: Arabidopsis amino acid transport protein I
C:Keyword: amino acid transport; transmembrane protein; transport protein

Query Match 100.0%; Score 2519; DB 2; Length 485;
Best Local Similarity 100.0%; Pred. No. 2.6e-196;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSFNTGHNHSTAESGDATVSDPTKRNVDGDERKRTGWTLTASAHITTAIVGSGVLSL 60
DB 1 MKSFNTGHNHSTAESGDATVSDPTKRNVDGDERKRTGWTLTASAHITTAIVGSGVLSL 60

QY 61 AMAIADGWTAGTISILLIFSFITFTSTMLADCYRAPDPVTGKNTYMDVVSYLGRK 120
DB 61 AMAIADGWTAGTISILLIFSFITFTSTMLADCYRAPDPVTGKNTYMDVVSYLGRK 120

QY 121 VOLGVAQYGNLIGTVGYTTTASISLVAVGKSNCFHDKHTADCTISNPYAVFGIIQ 180
DB 121 VOLGVAQYGNLIGTVGYTTTASISLVAVGKSNCFHDKHTADCTISNPYAVFGIIQ 180

QY 181 VIISOIPNFHKSFLSIMAAMVSTATTGIGLITVAVGKXKSMGTGAAGVDVTA 240
DB 181 VIISOIPNFHKSFLSIMAAMVSTATTGIGLITVAVGKXKSMGTGAAGVDVTA 240

QY 241 QKIMRSQAVGDIAPVAATVLEIQTFLRSSPAAENKMKRASLVGVSTTFEYILGCG 300

Db 241 QKIRSFQAVGDI AFAYATVLEIQDTRLSPAEKAMKASLGVSTTFFYLILCC 300
QY 301 IGYAAGNAPGDFLTDGFFFPFWLIDFANACIAVHLIGAYQVFAQIPQFQVEKCKNRN 360
Db 301 IGYAAGNAPGDFLTDGFFFPFWLIDFANACIAVHLIGAYQVFAQIPQFQVEKCKNRN 360
QY 361 YPDNKEITSEYSVNVDFLGFKNISLRLVWRTAYVVTWVAMIPFFFNAILGLIGAASF 420
Db 361 YPDNKEITSEYSVNVDFLGFKNISLRLVWRTAYVVTWVAMIPFFFNAILGLIGAASF 420
QY 421 WPLTVVFPVEMHIAQTKIKKYSARWIALKTMCVCLIVSLAAAGSIAGLISSVKTYKPF 480
Db 421 WPLTVVFPVEMHIAQTKIKKYSARWIALKTMCVCLIVSLAAAGSIAGLISSVKTYKPF 480
QY 481 RTMHE 485
Db 481 RTMHE 485
RESULT 2
T00620
probable amino acid transport protein T2711.3 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 22-Oct-1999
C:Accession: T00620
R:Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.; Araujo,
V.; Vyotskai, V.; Yu, G.; Ecker, J.; Theologis, A.; Davis, R.W.
submitted to the EMBL Data Library, September 1998
A:Reference number: Z14193
A:Accession: T00620
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-475 <PEP>
A:Cross-references: EMBL:AC004122; MID:g3176693; PIDN:AAC34329.1; PID:g3540179; GSPDB:GN
A:Experimental source: cultivar Columbia
C:Genetics:
A:Gene: ATSP:T2711.3
A:Map position: 1
A:introns: 108/1; 139/2; 211/1; 258/3; 334/3
C:Superfamily: Arabidopsis amino acid transport protein I
C:Keywords: amino acid transport; transmembrane protein; transport protein
Query Match 78.4%; Score 1974.5; DB 2; Length 475;
Best Local Similarity 75.4%; Pred. No. 3.3e-152;
Matches 364; Conservative 58; Mismatches 52; Indels 9; Gaps 2;
QY 1 MGSFNTGHNHSTAESGDAYTVSDPTKNVDEGKREKRTGTWLTASAHIIITAVIGSGVLSL 60
Db 1 MDAYN----NPSAVESGDA-----AVKSVDDGKREKRTGTWLTASAHIIITAVIGSGVLSL 51
QY 61 AWAIAGLWGIAGTSLILIFSTYFTSTMLADCYRAPDPVTGKRNYYTMDVVRSLVGRK 120
Db 52 AWAIAGLWVAGTTLVAFATITTYTSTLLADCYRSPDSITGTRNYYMGVVRSLVGRKK 111
QY 121 VOLCGVAQGNLIGVTVGTITATSLVAVGKSNCHDKGHTADCTISNYPWAVFGIITQ 180
Db 112 VOLCGVAQVNLVGVTVGTITATSLVAVGKSNCHDKGHTADCTISNYPWAVFGIITQ 171
QY 181 VILSOIPNPHKLSFLSILMAAVMSFTVATIGIGLAITAVAGGKGTMTGAVGVDVTA 240
Db 172 IILSQPNPHKLSFLSILAAVMSFTASIGIGLAITAVAGKIGKLTGTGTVGVDVTA 231
QY 241 QKIRSFQAVGDI AFAYATVLEIQDTRLSPAEKAMKASLGVSTTFFYLILCC 300
Db 232 EKVKWLFOAIGDIAFSAFTTILIEIQDTRLSPPEKKNVAVGVTTFVYLILCC 291
QY 301 IGYAAGNAPGDFLTDGFFFPFWLIDFANACIAVHLIGAYQVFAQIPQFQVEKCKNRN 360
Db 292 IGYAAGNAPGDFLTDGFFFPFWLIDFANACIAVHLIGAYQVFAQIPQFQVEKCKNK 351
QY 361 YPDNKEITSEYSVNVDFLGFKNISLRLVWRTAYVVTWVAMIPFFFNAILGLIGAASF 420

Db 352 WPOSFNINKEYSSKVPILGKCRVNLRLVWRTCYVVLVTFVAMIPFFFNAILGLGAFAP 411
QY 421 WPLTVVFPVEMHIAQTKIKKYSARWIALKTMCVCLIVSLAAAGSIAGLISSVKTYKPF 480
Db 412 WPLTVVFPVAMHIAQAKVKYSRRWALNLLVLVCLIVSALAAGVSIIGLINSVKYKPF 471
QY 481 RTM 483
Db 472 KNL 474
RESULT 3
T50691
amino acid permease 6 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 18-Jul-2001
C:Accession: T50691
R:Rentsch, D.; Hirner, B.; Schmelzer, E.; Frommer, W.B.
Plant Cell 8, 1437-1446, 1996
A:Title: Salt stress-induced proline transporters and salt stress-repressed broad specifi
A:Reference number: Z07208; MUID:96373203; PMID:8776904
A:Accession: T50691
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-481 <REN>
A:Cross-references: EMBL:X95736; PIDN:CAA65051.1
C:Genetics:
A:Gene: AAP6
C:Superfamily: Arabidopsis amino acid transport protein I
Query Match 74.3%; Score 1871; DB 2; Length 481;
Best Local Similarity 73.6%; Pred. No. 8.1e-144;
Matches 345; Conservative 60; Mismatches 62; Indels 2; Gaps 2;
QY 15 ESGDAYTVSDPTKNVDEGKREKRTGTWLTASAHIIITAVIGSGVLSLAWAIAGLWGIAGTS 74
Db 11 QSFPEHEIGDINKNFDEGDRDKRTGTWLTGSAHIIITAVIGSGVLSLAWAIAGLWVAGPA 70
QY 75 ILLPFSFTYFTSTMLADCYRAPDPVTGKRNYYTMDVVRSLVGRKVOLCGVAQGNLIG 134
Db 71 VLMAFSFTYFTSTMLADCYRSPDPVTGKRNYYTMEVVRSLVGRKVOLCGLAQYGNLIG 130
QY 135 VTVGVTTATSLVAVGKSNCFHDKGHTADCTISNYPWAVFGIITQVLSOIPNPHKLSF 194
Db 131 ITIGTITATSLVAVGKSNCFHKNHNVKCATSTNTPPMIIFAIITLSOIPNPHNLJSW 190
QY 195 LSILAAVMSFTVATIGIGLAITVA-GGKVGKTSMTGTVAGVDVTAQKIVRSFOAVGDI 253
Db 191 LSILAAVMSFCYASIGVLSIAKAAGGGEHVRTLTGTVTVIDYSGAEKIWRTFQAGDI 250
QY 254 APAYATVLEIQDTRLSS-PAENKAMKASLGVSTTTFYILCGCIGYAAFGNNAPG 312
Db 251 APAYATVLEIQDTRLKAGPPSENKAMKASLGVSTTTFYMLCGCGVAAAFGNDAFG 310
QY 313 DELTDFGFFPEFWLIDFANACIAVHLIGAYQVFAQIPQFQVEKCKNRNYPDKFITSYS 372
Db 311 NPLTFGFFPEFWLIDFANACIAVHLIGAYQVFCQIPQFQVESOSAKWPNKPFITGEYK 370
QY 373 VNVPLGKFNISLRLVWRTAYVVTWVAMIPFFFNAILGIGAASEFWPLTVVPPVEMH 432
Db 371 IHVPCGDFSNFLRLVWRTSYVVVAVVAMIPFFFNDFLGLIGNASFWPLTVVFPVEMH 430
QY 433 IAQTKIKKYSARWIALKTMCVCLIVSLAAAGSIAGLISSVKTYKPF 481
Db 431 IAQKIKPFSEFTWTLKILSWTCFIVSLVAAAGSVQGLIQSLKDPKPFQ 479
RESULT 4
T07130
probable amino acid transporter protein AAP1 - potato (fragment)
C:Species: Solanum tuberosum (potato)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
C:Accession: T07130

R.Kwart, M.

Submitted to the EMBL Data Library, November 1996

A:Reference number: Z15940

A:Accession: T07130

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-385 <KMA>

A:Cross-references: EMBL:Y09825; PIDN:CAA0968.1

A:Experimental source: cv. Desiree; tuber

C:Genetics:

A:Gene: AAP1

C:Superfamily: Arabidopsis amino acid transport protein I

C:Keywords: amino acid transport; transmembrane protein; transport protein

Query Match 63.1%; Score 1589.5; DB 2; Length 385;

Best Local Similarity 74.5%; Pred. No. 4e-121;

Matches 287; Conservative 55; Mismatches 42; Indels 1; Gaps 1;

QY 47 HITAVIGSVLSLMAIAQLGMIACTSIILIFSFITFTSTMLADCYRAPDPVTGKRY 106

Db 1 HITAVIGSVLSLMAIAQLGMIACTSIILIFSFITFTSTMLADCYRAPDPVTGKRY 60

QY 107 TMDVVRSLVGRKQVLCGVAOYGNLIGTVGTTTASISLVAVGKSNCFPHDKGHTADCT 166

Db 61 TMDVVRSLVGRKQVLCGVAOYGNLIGTVGTTTASISLVAVGKSNCFPHDKGHTADCT 120

QY 167 ISNYPMAVFGIIQVLSQIPNPKLSFLSISMAVMSFYATATIGLALATVAG-GKVGK 225

Db 121 ISNYPMAVFGIIQVLSQIPNPKLSFLSISMAVMSFYATATIGLALATVAG-GKVGK 180

QY 226 TSMGTAVGVDTAAQKIRSFQAVGDIAPAYAVATVLEIOTLRSSPAENKMKRASL 285

Db 181 TALTGVVGVDSVGTETKVRSPQAIQDIAFAVYSVTLIEIOTLRSSPAENKMKRASL 240

QY 286 VGVSTTTPFYIICGCTGVAFCGNAGDFLTDFGFEPPFWLIDFANACIAVHLIGAYQVF 345

Db 241 VGVSTTTPFYIICGCTGVAFCGNAGDFLTDFGFEPPFWLIDFANACIAVHLIGAYQVF 300

QY 346 AOPFQFVEKCKGNRNPDKRFTSESVNVVPLFGKFNISLFLWRATVAVVITTVAMIF 405

Db 301 COPLVGFVGRGSRERPDKFTISEYAMQVPMCGITYNLMLFLWRATVAVVITTVAMIF 360

QY 406 PPFNAIILGIGAEMWPLTVPEVE 430

Db 361 PPFNDELIGIGAEMWPLTVPEVE 385

RESULT 5

T10100 amino acid transport protein AAP1 - castor bean

N:Alternate names: amino acid permease; amino acid transporter

C:Species: Ricinus communis (castor bean)

C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C:Accession: T10100

R:Marlier, A.C.; Neelam, A.; Bick, J.A.; Hall, J.L.; Williams, L.E.

Biochim. Biophys. Acta 1373, 321-331, 1998

A:Title: Cloning of a cDNA coding for an amino acid carrier from Ricinus communis (RCAAF

A:Reference number: Z16947; MUID:98405925; PMID:973391

A:Accession: T10100

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-486 <MAR>

A:Cross-references: EMBL:AJ007574; NID:g3293030; PIDN:CAA0563.1; PID:g3293031

A:Experimental source: cultivar sanguineous

C:Genetics:

A:Gene: AAP1

C:Function:

A:Description: transport of basic amino acids e.g. histidine

C:Superfamily: Arabidopsis amino acid transport protein I

C:Keywords: amino acid transport; transmembrane protein; transport protein

Query Match 61.2%; Score 1540.5; DB 2; Length 486;

Best Local Similarity 58.2%; Pred. No. 4.9e-117;

Matches 285; Conservative 82; Mismatches 106; Indels 17; Gaps 2;

QY 5 NTEGHNSHSTAESGDAYTVSDPTKN-----VDEGREGKRTGTWLTASAHITAVIG 54

Db 4 NTAAGNH-----PHQPVDSINNQTQVSGKMLDDDRKTRGTGTWLTASAHITAVIG 56

QY 55 SGVLSLMAIAQLGMIACTSIILIFSFITFTSTMLADCYRAPDPVTGKRYTMDVRS 114

Db 57 SGVLSLMAIAQLGMIACTSIILIFSFITFTSTMLADCYRAPDPVTGKRYTMDVRS 116

QY 115 YLGRKQVLCGVAOYGNLIGTVGTTTASISLVAVGKSNCFPHDKGHTADCTISNYPMA 174

Db 117 YLGRKQVLCGVAOYGNLIGTVGTTTASISLVAVGKSNCFPHDKGHTADCTISNYPMA 176

QY 175 VGGIIQVLSQIPNPKLSFLSISMAVMSFYATATIGLALATVAG-GKVGK 234

Db 177 VGGIIQVLSQIPNPKLSFLSISMAVMSFYATATIGLALATVAG-GKVGK 236

QY 235 VDTAAQKIRSFQAVGDIAPAYAVATVLEIOTLRSSPAENKMKRASLGVSTTTF 294

Db 237 VDTAAQKIRSFQAVGDIAPAYAVATVLEIOTLRSSPAENKMKRASLGVSTTTF 296

QY 295 YILCGCTGVAFCGNAGDFLTDFGFEPPFWLIDFANACIAVHLIGAYQVFQFVE 354

Db 297 YILCGCTGVAFCGNAGDFLTDFGFEPPFWLIDFANACIAVHLIGAYQVFQFVE 356

QY 355 KCKNNRNPDKRFTSESVNVVPLFGKFNISLFLWRATVAVVITTVAMIFPFENALIG 414

Db 357 KCKNNRNPDKRFTSESVNVVPLFGKFNISLFLWRATVAVVITTVAMIFPFENALIG 416

QY 415 IGASFWPLTVFPEVEMHIAQTKIKKYSARWIALKTMCVCLIVSLAAGSIAGLISSV 474

Db 417 IGASFWPLTVFPEVEMHIAQTKIKKYSARWIALKTMCVCLIVSLAAGSIAGLISSV 476

QY 475 KTYKPEPTMH 484

Db 477 KTYKPEPTMH 486

RESULT 6

H96802 probable amino acid carrier [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: H96802

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federpiel, N.A.; Keul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.-J.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,

Rizzio, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: H96802

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-476 <STO>

A:Cross-references: GB:AE005173; NID:g11079491; PIDN:AG29203.1; GSPDB:GN00141

C:Genetics:

A:Gene: P2P24.9

A:Map position: 1

C:Superfamily: Arabidopsis amino acid transport protein I

Query Match 59.7%; Score 1504; DB 2; Length 476;

Best Local Similarity 57.6%; Pred. No. 4.3e-114;

Matches 273; Conservative 97; Mismatches 100; Indels 4; Gaps 4;

QY 10 NHSTAESGDAYTVSDPTKNVDEGREGKRTGTWLTASAHITAVIGSVLSLMAIAQLGM 69

Db 4 NHQTVLAVD-MPQTGSKYLDGKVKRGTGVSMTASAHITAVIGSVLSLMAIAQLGM 62

197 GGLGVSKVENKEIKSGITGVITLSTGVTSQKWTPTQSLGNIAFAYSYSMIII 256
 QY 265 EIOTLRSSPAENKAMKASLWGVSTTFFYILCGCIGAAFGNNAFGDITDFGFEFP 324
 Db 257 EIQDVKRPPAEVNMTRKATPFVSAVTTFYMLCGCVGAAGDANFGSLAHGGRNRY 316
 QY 325 WLIDFANACIAVHLIGAYOVFAQPIFQVEKKCNKNYPDNKEITSEYSVNVPFLK-FNI 383
 Db 317 WLDDIANIAIVHLIGAYOVQCPLFAFVEKESRRFPSEFVTEKEIKIQL-FPGKPFNL 375
 QY 384 SLFRLVMTATVYVITTVVAMIFPPFNALIGLGAASFWPLTYFPFPMIAIQTAKKYYSA 443
 Db 376 NLFRLVMTATVYVITTVVAMIFPPFNALIGLGAASFWPLTYFPFPMIAIQTAKKYYSA 435
 QY 444 RMIALKTCVCLIVSLAAGSIAGLISSVKTYKPFRT 482
 Db 436 KMWCLQVLSVTCLEFVSVAAAGSVIGIVSDLVKTYKPFQS 474

RESULT 9

T07131
 probable amino acid transporter protein AAP2 - potato (fragment)
 C:Species: Solanum tuberosum (potato)
 C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 20-Jun-2000
 C:Accession: T07131
 R:Kwart, M.
 submitted to the EMBL Data Library, November 1996
 A:Accession: Z15940
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-376 <KMA>
 A:Cross-references: EMBL:Y09826; PIDN:CAA70969.1
 A:Experimental source: cv. Desiree, leaf
 C:Genetics:
 A:Gene: AAP2
 C:Superfamily: Arabidopsis amino acid transporter protein I
 C:Keywords: amino acid transport; transmembrane protein; transport protein

Query Match 52.6%; Score 1326; DB 2; Length 376;
 Best Local Similarity 62.9%; Pred. No. 8.7e-100;

Matches 237; Conservative 68; Mismatches 70; Indels 2; Gaps 2;

QY 47 HITTAIVIGSGVSLAMALQGLMAGTSLILFSTFTSTMLADCPAPDPVTGKRY 106
 Db 1 HITTAIVIGSGVSLAMATQGLMAGPTVLLFFVTYTSALSDCYRTGDPVTGKRY 60
 QY 107 TYMDVRSYLGKRVQLCGVAQYGNLIGVTVYTTASISLVAVGKSCFHDKHTADCT 166
 Db 61 TYMDAVRANLGGFQYKICGVIQVANFGVAIGYTIASISMAVNRNCFHKQGHRAACN 120
 QY 167 ISNYPMAVFGIIQVILSQIPNFKLSFSLMAVMSFYATITGIGLAIATYA-GKRVK 225
 Db 121 VASTPPIIFGWEIIFSDIPDFDQISWLSIYAAVMSFYSTTIGLGAOVAETGKI-E 179
 QY 226 TSMGTGAVGVDTAAOKTMRSPQAVGDIAPAYAVATVLEIOTLRSSPAENKAMKASL 285
 Db 180 GSLTGISTEVTETWKMRSPQALAIAPAYSILILEIOTLRSSPAENKAMKASL 239
 QY 286 VGVSTTFFYILCGCIGAAFGNNAFGDITDFGFEFPMLIDFANACIAVHLIGAYOVF 345
 Db 240 ISVAATVTFYMLCGCGVAAFGDQSPGNLITGFGFYNPWLIDIANVALVHLVGAYOVY 299
 QY 346 AQPIQFQVFKCNKNYPDNKEITSEYSVNVPFLGKNISLFLVMTATVYVITTVVAMIF 405
 Db 300 CQPLFAFVFKTATWEPDKIITKEIDVDPGKPKMLFRLVMTATVYVITTVISM 359
 QY 406 PPFNALLIGLGAASFWP 422
 Db 360 PPFNDVVGILGAFGFWP 376

RESULT 10

T09840
 amino acid transporter protein AAP2 - castor bean (fragment)
 N:Alternate names: amino acid permease; amino acid transporter
 C:Species: Ricinus communis (castor bean)
 C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000
 C:Accession: T09840
 R:Blück, J.A.; Neelam, A.; Hall, J.L.; Williams, L.E.
 Plant Mol. Biol. 36, 377-385, 1998
 A:Title: Amino acid carriers of Ricinus communis expressed during seedling development:
 A:Reference number: Z16878; MUID:98145472; PMID:9484478
 A:Accession: T09840
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-284 <BLC>
 A:Cross-references: EMBL:Y11121; NID:G1839021; PIDN:CAA72006.1; PID:G1839022
 A:Experimental source: strain Sanguineas; tissue-type root; clone RcAAP2
 C:Superfamily: Arabidopsis amino acid transporter protein I
 C:Keywords: amino acid transport; transmembrane protein; transport protein

Query Match 46.2%; Score 1163; DB 2; Length 284;
 Best Local Similarity 75.0%; Pred. No. 1e-66;

Matches 213; Conservative 37; Mismatches 34; Indels 0; Gaps 0;

QY 63 AIAQGLMAGTSLILFSTFTSTMLADCPAPDPVTGKRYTYMDVRSYLGKRVQ 122
 Db 1 AIAQGLMAGTSLILFSTFTSTMLADCPAPDPVTGKRYTYMDVRSYLGKRVQ 60
 QY 123 LCGVAQYGNLIGVTVYTTASISLVAVGKSCFHDKHTADCTISNYPMAVFGIIQV 182
 Db 61 FCGISQYANLVGIVTVYTTASISMAVNRNCFHRHGAACHTSNNYMTIFACTIQL 120
 QY 183 LSQIPNFKLSFSLMAVMSFYATITGIGLAIATVAGKRVKTSMTGTAVGVDTAAOK 242
 Db 121 LSQIPNFKLSFSLMAVMSFYATITGIGLAIATVAGKRVKTSMTGTAVGVDTAAOK 180
 QY 243 IWRSPQAVGDIAPAYAVATVLEIOTLRSSPAENKAMKASLWGVSTTFFYILCGCIG 302
 Db 181 IWRSPQISGDIAPAYAVATVLEIOTLRSSPAENKAMKASLWGVSTTFFYILCGCIG 240
 QY 303 YAAFGNNAFGDITDFGFEFPMLIDFANACIAVHLIGAYOVFA 346
 Db 241 YAAFGNNAFGDITDFGFEFPMLIDIANVALVHLIGAYOVFS 284

RESULT 11

A57479
 amino acid transporter protein AAP3 - Arabidopsis thaliana
 N:Alternate names: amino acid permease; amino acid transporter
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 20-Jun-2000
 C:Accession: A57479; S51168
 R:Fischer, W.N.; Kwart, M.; Hummel, S.; Frommer, W.B.
 J. Biol. Chem. 270, 16315-16320, 1995
 A:Title: Substrate specificity and expression profile of amino acid transporters (AAPs)
 A:Reference number: A57479; MUID:95332342; PMID:7608199
 A:Accession: A57479
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-405 <FIS>
 A:Cross-references: GB:X77499
 R:Fischer, W.N.
 submitted to the EMBL Data Library, January 1994
 A:Reference number: S51168
 A:Accession: S51168
 A:Molecule type: mRNA
 A:Residues: 1-21, 'U', '23-60, 'GWLAVPVLMLFSA', '61-136, 'KSGGKDBCHMNSNP', '137-205, 'IWRFGALGDI
 A:Cross-references: EMBL:X77499; NID:g608668; PID:g608669
 C:Genetics:
 A:Gene: AAP3
 C:Superfamily: Arabidopsis amino acid transporter protein I
 C:Keywords: amino acid transport; transport protein

Query Match 45.8%; Score 1153.5; DB 2; Length 405;

```

Db      55  STLLSCYRTGDPVSGKRNYYTMDAYRSILGGFRFKICGLIOYLNLFGITVGYTIAASTS 114
QY      147  LVAGKSNCFDKGHTADCTISNYPMAVFGIIQVILSQIPNFHKLKSLFSLMAAVMSFTY 206
Db      115  MMAIKSNCFH-----YMIWFGYTEILLSSQIKQFDQIWLWLSLVAIIMSFTY 160
QY      207  ATTIGLAIATVAGGKVGKTSMTGTAVGVDVTAQAOKIMRSFOAVGDIAPAYATVLI 266
Db      161  SAIGLAGLIIQVAANGVWKGSLTGISGA-VTQTK-----YSVSVVLLIEI 205
QY      267  QDTLRSPAEKAMKASLASLVGVSTTTFFVILCGCTGYAAGFNNAPOGDFLTDFGPEPFWL 326
Db      206  QDVRSPPAESKTMKIAATRIASIAVTTTFMYLGCQMGYAAFGDKAPGNLLTGFEYVFWL 265
QY      327  IDPANACIAVHLIGAYQVPAQPIFQVEKKRNRYDNPKFIITSEXSVMYP-FLGKFNISL 385
Db      266  LDVAN-----AQPIFAFIEKQLAARPDSDLVTKVEYIRIPGRSPSYKVN 311
QY      386  FRLVMTAYVYVTVTVAMIFPFENALIGLIGAASFPLTVYFVPMHIAQTAKIKKYSARW 445
Db      312  FRAVYRSGFVLLTVISMLMPFENDVVGILGALGF-----RKVERMSMKW 357
QY      446  IALKTWYCVCLIVSLLAAGASAGLISSVKTKPFRPTMH 484
Db      358  VCLQMLSCGLMITLVAGVGSIAGVMLDLKVKYKPKTTY 396

RESULT 13
C57479
A:Name: amino acid transport protein AAP5 - Arabidopsis thaliana
N:Alternate names: amino acid permease; amino acid transporter
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Jan-1996 #sequence_revision 19-Jan-1996 #text_change 20-Jun-2000
C:Accession: C57479; S51170
R:Fischer, W.N.; Kwart, M.; Hummel, S.; Frommer, W.B.
J. Biol. Chem. 270, 16315-16320, 1995
A:Title: Substrate specificity and expression profile of amino acid transp
A:Reference number: A57479; MUID:95332342; PMID:7608199
A:Accession: C57479
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-411 <PIS>
A:Cross-references: GB:X77501
R:Fischer, W.N.
submitted to the EMBL Data Library, January 1994
A:Reference number: S51168
A:Accession: S51170
A:Molecule type: mRNA
A:Residues: 1-58, <GWIGGPVAMLLFSF>, 59-88, '1', 90-134, <MGNPNDPCHVGNV>, 135-211
A:Cross-references: EMBL:X77501; NID:9608672; PIDN:CAA54632.1; PID:9608673
C:Genetics:
A:Gene: AAP5
C:Superfamily: Arabidopsis amino acid transport protein I
C:Keywords: amino acid transport; transport protein

Query Match 44.0%; Score 1109.5; DB 2; Length 411;
Best Local Similarity 47.1%; Pred. No. 3.3e-82;
Matches 216; Conservative 80; Mismatches 86; Indels 77; Gaps

QY      30  DEDGRKRTGTWLTASAHITAVIGSGVLSLAWAQLGWIAGTSTILFFSFTYPTSTM 89
Db      21  DDDGRPKRTGTWLTASAHITAVIGSGVLSLAWAQAQI-----VTFVTSTL 66
QY      90  LADCYRAPDPVTKRKNYYTMDVVRSLGGRKQVOLCSVAOYGNLIGVTVGYTTASISLVA 149
Db      67  LCSYRSGDSVTGKRNYYTMDATHSNLGGIKVKYCVGWQYVNLFGTAIGYTTASISLVA 126
QY      150  VGKSNCFDKGHTADCTISNYPMAVFGIIQVILSQIPNFHKLKSLFSLMAAVMSFTYAT 209
Db      127  IQRTSC-----QQMTAFGIQVIFSQIPDFDQLWLSLVAIIMSFTYAT 172
QY      210  GIGLAIATVAGGKVGKTSMTGTAVG-----DVTAQKIKWRSFQAVGDIAPAYATVLI 264

```

Db 173 GGLGSGSKVENKEIKSGTGTGVTLSGVTSSOK-----YSYSMLI 218
 QY 265 EIQDELSSPAENKMKRSLVGVSTTTFFYILGCGIGAFGNNAQDFLDFGPEEP 324
 Db 219 EIQDTKSPAEVNMKRTFTFSVAATTFYMLCGCVGAAGDANPGLMLHAGGRNRY 278
 QY 325 WLIDPANACIAVHLIGAYQVFAQPIFOFYEKKCNRNYPDNKFTSESVNVPLGR-FNI 383
 Db 279 WLIDIAN-----LQPLFAFEKEKSRFRPESEFVTKIKIQL-FPGKPFNL 323
 QY 384 SLFRLVWRVAVVITTVAMIFPPFNAIIGLIGASFWLTYFPVEMHIAQTKIKYGA 443
 Db 324 NLFRLVWRVFFVITTLIMLMEFFNDVVGGLGALGFW-----KNVPRMGV 369
 QY 444 RWIALKTCYCVIIVSLAAGSIAGLISVKTYPKPT 482
 Db 370 KWCICQVLSVTCIFVSVAAAGSVIGIVSDIKYKPFQS 408

RESULT 14

T09843

amino acid transport protein AAP1 - caenor bean (fragment)

N/Alternate names: amino acid permease; amino acid transporter

C/Species: Ricinus communis (castor bean)

C/Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 21-Jul-2000

C/Accession: T09843

R/Bick, J.A.; Neelam, A.; Hall, J.L.; Williams, L.E.

Plant Mol. Biol. 36, 377-385, 1998

A/Title: Amino acid carriers of Ricinus communis expressed during seedling development:

A/Reference number: Z16878; PMID:98145472; PMID:9484478

A/Accession: T09843

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-284 <BIC>

A/Cross-references: EMBL:Z68759; NID:G2995320; PIDN:CAA32992.1; PID:G2995321

A/Experimental source: strawn Sanguineas; tissue-type root; clone RCAAP1

C/Superfamily: Arabidopsis amino acid transport protein I

C/Keywords: amino acid transport; transmembrane protein; transport protein

Query Match 38.3%; Score 966; DB 2; Length 284;

Best Local Similarity 61.3%; Pred. No. 9.2e-71;

Matches 174; Conservative 50; Mismatches 60; Indels 0; Gaps 0;

QY 63 ATAQGLMGITSTLIFSPITVFTSTMLADCYRAPDPVTKRNYTMDVRSYLGARKYQ 122
 Db 1 ATAQGLMGIPAVMFLFSLVTTTSTLSCHKSGDPVNGKRYTMDAVRTNLGSAKAK 60
 QY 123 LCGVAQGNLIGTVGTTTASISLVAVGKSCFHDGHTADCTISNYPMAVEGIIQYI 182
 Db 61 LCGFVQYLLFGVALIGYTITASSISMAIKRSCFHKSGGKPNCHIANPFIAMFGIAETI 120
 QY 183 LSGIENFHLKSLSTMAAVMSFTYATIGLAITATYAGGKVGKTSMTGTAIVGVDTAAQ 242
 Db 121 FSGIPEFDLMLWISIAAVMSFTYATIGLAITATYAGGKVGKTSMTGTAIVGVDTAAQ 180
 QY 243 ITRSPAVGDIAFAVAVATVLEIOTLRSSPAENKMRASVGVSTTFFYILGCGIC 302
 Db 181 ITRSPALDIDIAAAYSITLLEIOTLRSSPESKTKKATLISVAVTTLFFMLGCGIC 240
 QY 303 YAAFGNNAQDFLTDGFEFEPFLIDPANACIAVHLIGAYQVFA 346
 Db 241 YAAFGMSFGNLTGRGFNFPYLLDIANVAIVHLIGAYQVFA 284

RESULT 15

C86378

protein F219.6 [imported] - Arabidopsis thaliana

C/Species: Arabidopsis thaliana (mouse-ear cross)

C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001

C/Accession: C86378

R/Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A/Authors: Salberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Frazer, C.M.; Venter, J.C.; Davis, R.W.

A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A/Reference number: A6141; PMID:21016719; PMID:11130712

A/Accession: C86378

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-441 <STO>

A/Cross-references: GB:AB005172; NID:G9743356; PIDN:AAF97980.1; GSPDB:GN00141

C/Genetics:

A/Map position: 1

C/Superfamily: Arabidopsis amino acid transport protein I

Query Match 22.1%; Score 557; DB 2; Length 441;

Best Local Similarity 31.1%; Pred. No. 2.1e-37;

Matches 145; Conservative 80; Mismatches 187; Indels 54; Gaps 12;

QY 27 KAVDE--DGRKRTGTMLTASAHITTAIVGSGVLSLAMAIAQLGWTAGTSLILISFTY 84
 Db 17 KAVDDMLPTSSRNKAKWMSAFHNVTAMVAGAVLSIPYMSMLGMPGVITIMWSWITL 76
 QY 85 FSTMLADCYRAPDPVTKRNTVMDVRSYLGGRKQVLCGVAQYGNL--IGTVGTTT 142
 Db 77 YTLQVQVNH--ELVPGRLDRHIE--LQGHAFGEGLGMIYVPOQLIVEGVDIYVMT 132
 QY 143 ASISLVAVGKSCFHDGHTADC--TISNYPMAVEGIIQYIISQIPNFHKLSTLSMAAV 201
 Db 133 GGASLKKVHQLVC-----PCKEIRTFMIMIFASVHFVSHLPNFSISITSLAAAV 185
 QY 202 MSFTTATIGLAITAVAGGKVGKTSMTGTAIVGVDTAAQIKWSFQAVGDIAPAYAVAT 261
 Db 186 MSFTYSTIAMAASVHGVHPDVSPRASTDVG-----KVFNFALAGDVAFAYAGHN 238
 QY 262 VLEIOTLRSSP--AENKMRASVGVSTTFFYILGCGICGYAFAFGNNAQDFLTDG 319
 Db 239 VLEIQTATPSTPEMPSKVPWNRGVIVAVIICFPVAFIYIYFGNSVDNIIITLE 298
 QY 320 FPEPFLIDPANACIAVHLIGAYQVFAQPIFOFVE-----KKCRNYPDNKFTSESVNV 375
 Db 299 --KPIWLMAMNFVYIHIGSYQIFAMPVFDMLEVLYKGN----- 339
 QY 376 PFLGKFNISL--FRLVWRVAVVITTVAMIIFPFNAIIGLIGASFWLTYFPVEMHIA 434
 Db 340 -----FNPFSKLRFTIRSLYVAFMTVAICVPEFGGLGFGGFAFAPTTYLLPCIMLV 394
 QY 435 QTKIKYKSARWIALKTCYCVIIVSLAAGSIAGLISVKTYPK 480
 Db 395 LKKPRKFGLSWTANWFCIIIVGLITLAPIGGLRTIINAKTYKPF 440

Search completed: November 22, 2002, 13:30:45

Job time : 23 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2002, 13:28:57 ; Search time 35 Seconds

(without alignments)
2855.224 Million cell updates/sec

Title: US-09-854-562-2

Perfect score: 2519
Sequence: 1 MKSFNTGHNHSTAESGDY.....SIAGLISSVKYKPPRTMHE 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2519	100.0	485	10	Q42400 arabidopsis
2	2510	99.6	485	10	Q48546 arabidopsis
3	1974.5	78.4	475	10	O80592 arabidopsis
4	1871	74.3	481	10	P92934 arabidopsis
5	1589.5	63.1	385	10	P93561 solanum tub
6	1540.5	61.2	486	10	O82044 ricinus com
7	1504	59.7	476	10	O39134 arabidopsis
8	1501	59.6	493	10	O38967 arabidopsis
9	1496	59.4	466	10	O8RWAB arabidopsis
10	1495	59.3	466	10	O9FN04 arabidopsis
11	1494	59.3	466	10	O39135 arabidopsis
12	1484.5	58.9	379	10	O92PM7 nepenthes a
13	1471.5	58.4	488	10	O8RZP7 oryza sativ
14	1462	58.0	481	10	O93X13 vicia faba
15	1442.5	57.3	466	10	O92RS1 ricinus com
16	1441.5	57.2	513	10	O9ARG2 glycine max

17	1427.5	56.7	480	10	Q39136 arabidopsis
18	1427.5	56.7	509	10	O92R62 vicia faba
19	1417.5	56.3	475	10	O93X15 vicia faba
20	1406.5	55.8	476	10	O9C6V2 arabidopsis
21	1326	52.6	376	10	P93562 solanum tub
22	1294.5	51.4	377	10	O92PM6 nepenthes a
23	1285	51.0	486	10	O93X14 vicia faba
24	1200.5	47.7	376	10	O92PM5 nepenthes a
25	1198.5	47.6	469	10	O8S3N9 oryza sativ
26	1179	46.8	467	10	O9FP99 arabidopsis
27	1163	46.2	284	10	P93506 ricinus com
28	1016.5	40.4	263	10	O9SEM7 vicia faba
29	966	38.3	284	10	O41132 ricinus com
30	860.5	34.2	261	10	O9SEM6 vicia faba
31	840.5	33.4	303	10	O941B6 arabidopsis
32	744.5	29.6	259	10	O9SEM5 vicia faba
33	557	22.1	441	10	O91RB5 arabidopsis
34	548	21.8	441	10	O9SR44 arabidopsis
35	518	20.6	154	10	O8VWU1 arabidopsis
36	514	20.4	161	10	O8VWU9 narcissus p
37	507.5	20.1	462	10	O40414 nicotiana s
38	501	19.9	446	10	O9FKS8 arabidopsis
39	496	19.7	440	10	O9C6M2 arabidopsis
40	493	19.6	450	10	O9M9H9 arabidopsis
41	485	19.3	809	10	O9C7J3 arabidopsis
42	461	18.3	434	10	O9C9J0 arabidopsis
43	437	17.3	446	10	O24405 arabidopsis
44	436.5	17.3	455	10	O9SGJ3 arabidopsis
45	436.5	17.3	479	10	O9SS86 arabidopsis

ALIGNMENTS

RESULT 1	ID	Q42400	PRELIMINARY;	PRT;	485 AA.
AC	Q42400;				
DT	01-NOV-1996	(TREMBLrel. 01, Created)			
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)			
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)			
DE	Amino acid permease 1 (F19C14.3 protein).				
GN	AAPI OR F19C14.3 OR ZCF54.				
OS	Arabidopsis thaliana (Mouse-ear cress).				
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;				
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;				
OC	eurosid II; Brassicales; Brassicaceae; Arabidopsis.				
OX	NCBI_TaxID=3702;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=SSP. LANDSBERG;				
RX	MEDLINE=93317600; PubMed=8327465;				
RA	Frommer W.B., Hummel S., Riesmeier J.W.,				
RT	"Expression cloning in yeast of a cDNA encoding a broad specificity				
RT	amino acid permease from Arabidopsis thaliana."				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:5944-5948(1993).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=93361464; PubMed=8356039;				
RA	Hsu L.C., Chiou T.J., Chen L., Bush D.R.,				
RT	"Cloning a plant amino acid transporter by functional complementation				
RT	of a yeast amino acid transport mutant."				
RL	Proc. Natl. Acad. Sci. U.S.A. 90:7441-7445(1993).				
RN	[3]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CV. COLUMBIA;				
RA	Liu S.X., Sakano H., Yu G., Lee J.M., Ienz C., Toriumi M.,				
RA	Chin C., Chiou J., Choi E., Chung M., Gonzalez A., Hwang B., Liu A.,				
RA	Vaynberg M., Alfati H., Brooks S., Buehler E., Chao Q., Conn L.,				
RA	Conaway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,				
RA	Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,				
RA	Ecker J.R., Federle N.A., Theologis A.;				
RT	"The sequence of BAC F19C14 from Arabidopsis thaliana chromosome 1."				

Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.

[4]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Kato A., Kameda Y.;
RT "Long repeat sequence within a genomic region located around the 100 map unit of chromosome 1 in Arabidopsis thaliana.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20018182; PubMed=10548732;
RA Kato A., Suzuki M., Kuwahara A., Ooe H., H-Inaba K., Kameda Y.;
RT "Isolation and analysis of a cDNA within a 300 kb Arabidopsis thaliana genomic region located around the 100 map unit of chromosome 1.";
RL Gene 239:309-316(1999).
RN [6]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=20263341; PubMed=10805591;
RA Suzuki M., Kato A., Kameda Y.;
RT "An RNA-Binding protein, ACRBP1, is expressed in actively proliferative regions in Arabidopsis thaliana.";
RL Plant Cell Physiol. 41:282-288(2000).
RN [7]
RN SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Suzuki M., Kato A., Nagata N., Kameda Y.;
RT "An xylanase, AtXyl1, is predominantly expressed in vascular bundles, and two putative xylanase genes were identified in the Arabidopsis thaliana genome.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; X67124; CAA47603.1; -;
DR EMBL; L16240; AAA32726.1; -;
DR EMBL; AC08051; AAF82252.1; -;
DR EMBL; AB077822; BAB83868.1; -;
DR InterPro; IPR002422; AA/rel_prmase2.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF01490; Aa_trans; 1.
DR PROSITE; PS00063; ALDOXETO_REDUCTASE_3; UNKNOWN 1.
SQ SEQUENCE 485 AA; 52895 MW; 722E3895937CCC5B CRC64;

Query Match 100.0%; Score 2519; DB 10; Length 485;
Best Local Similarity 100.0%; Pred. No. 2.2e-188; Mismatches 0; Indels 0; Gaps 0;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSFNTEGHNHSTAESGDAYTVSDPTKNVDEDEGREKRTGTWLTASAHITAVIGSGVLSL 60
DB 1 MKSFNTEGHNHSTAESGDAYTVSDPTKNVDEDEGREKRTGTWLTASAHITAVIGSGVLSL 60

QY 61 AWAIQLGWIAGTSILLIFSPITYFTSTMLADCYRAPDPVTGKRNYYMDVVRSYLGGK 120
DB 61 AWAIQLGWIAGTSILLIFSPITYFTSTMLADCYRAPDPVTGKRNYYMDVVRSYLGGK 120

QY 121 VOLCGVAQYGNLIGVTGVTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 180
DB 121 VOLCGVAQYGNLIGVTGVTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 180

QY 181 VILSQIPNPHKLSFLSILMAAVMSFTYATIGLAIATVAGGKVGKTSMTGTAVGVDVTA 240
DB 181 VILSQIPNPHKLSFLSILMAAVMSFTYATIGLAIATVAGGKVGKTSMTGTAVGVDVTA 240

QY 241 QKIWRSFOAVGDIAPAYATVLIETQDTRSSPAENKAMKASLVGVSTTTFFYLGC 300
DB 241 QKIWRSFOAVGDIAPAYATVLIETQDTRSSPAENKAMKASLVGVSTTTFFYLGC 300

QY 301 IGVAAGNNAPGDFLTDGFFEPFWLIDFANACIAVHLIGAYQVPAQPIFQVVEKKNRN 360
DB 301 IGVAAGNNAPGDFLTDGFFEPFWLIDFANACIAVHLIGAYQVPAQPIFQVVEKKNRN 360

QY 361 YPDNKEITSEYSVNPFLGKFNISLRLVWRTAVYVITTVVAMIFFPFAILGLIGAASF 420
DB 361 YPDNKEITSEYSVNPFLGKFNISLRLVWRTAVYVITTVVAMIFFPFAILGLIGAASF 420

QY 421 WPLTVYFPVEMHIAQTJKIKYSARWIAKTCYVCLIVSLAAAGSIAGLISSVKTYPK 480
DB 421 WPLTVYFPVEMHIAQTJKIKYSARWIAKTCYVCLIVSLAAAGSIAGLISSVKTYPK 480

QY 481 RTMHE 485
DB 481 RTMHE 485

RESULT 2
O48546 PRELIMINARY; PRT; 485 AA.
AC O48546;
DT 01-JUN-1998 (TRENBLrel. 06, Created)
DT 01-JUN-1998 (TRENBLrel. 06, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE Neutral amino acid transport system II.
GN NAT2.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsie.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX MEDLINE=93361464; PubMed=8356039;
RA Hsu L.-C., Chiou T.-J., Chen L., Bush D.R.;
RT "Cloning a plant amino acid transporter by functional complementation of a yeast amino acid transport mutant.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:7441-7445(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RX Chang H.-C., Bush D.R.;
RT "Genomic Sequence of NAT2/AAPI of Arabidopsis";
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF031649; AAB87674.1; -;
DR InterPro; IPR002422; AA/rel_prmase2.
DR InterPro; IPR001395; Aldo/ket_red.
DR Pfam; PF01490; Aa_trans; 1.
DR PROSITE; PS00063; ALDOXETO_REDUCTASE_3; UNKNOWN 1.
SQ SEQUENCE 485 AA; 52943 MW; C4F3C49A538FC06B CRC64;

Query Match 99.6%; Score 2510; DB 10; Length 485;
Best Local Similarity 99.8%; Pred. No. 1.1e-187; Mismatches 0; Indels 1; Gaps 0;
Matches 484; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MKSFNTEGHNHSTAESGDAYTVSDPTKNVDEDEGREKRTGTWLTASAHITAVIGSGVLSL 60
DB 1 MKSFNTEGHNHSTAESGDAYTVSDPTKNVDEDEGREKRTGTWLTASAHITAVIGSGVLSL 60

QY 61 AWAIQLGWIAGTSILLIFSPITYFTSTMLADCYRAPDPVTGKRNYYMDVVRSYLGGK 120
DB 61 AWAIQLGWIAGTSILLIFSPITYFTSTMLADCYRAPDPVTGKRNYYMDVVRSYLGGK 120

QY 121 VOLCGVAQYGNLIGVTGVTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 180
DB 121 VOLCGVAQYGNLIGVTGVTITASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 180

QY 181 VILSQIPNPHKLSFLSILMAAVMSFTYATIGLAIATVAGGKVGKTSMTGTAVGVDVTA 240
DB 181 VILSQIPNPHKLSFLSILMAAVMSFTYATIGLAIATVAGGKVGKTSMTGTAVGVDVTA 240

QY 241 QKIWRSFOAVGDIAPAYATVLIETQDTRSSPAENKAMKASLVGVSTTTFFYLGC 300
DB 241 QKIWRSFOAVGDIAPAYATVLIETQDTRSSPAENKAMKASLVGVSTTTFFYLGC 300

QY 301 IGVAAGNNAPGDFLTDGFFEPFWLIDFANACIAVHLIGAYQVPAQPIFQVVEKKNRN 360
DB 301 IGVAAGNNAPGDFLTDGFFEPFWLIDFANACIAVHLIGAYQVPAQPIFQVVEKKNRN 360

QY 361 YPDNKFITSESVNVPFLCKENISLFLVWRTAYVITTVVAMIPFPFNALIGLGAASF 420
DB 361 YPDNKFITSESVNVPFLCKENISLFLVWRTAYVITTVVAMIPFPFNALIGLGAASF 420
QY 421 WPLTYVPFVEMHIAOTKIKKYSARWIALKTMCVCLIVSLAAGSIAGLISSVTKYKPF 480
DB 421 WPLTYVPFVEMHIAOTKIKKYSARWIALKTMCVCLIVSLAAGSIAGLISSVTKYKPF 480
QY 481 RTMHE 485
DB 481 RTMHE 485
RESULT 3
ID 080592 PRELIMINARY; PRT: 475 AA.
AC 080592
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Putative amino acid permease.
GN T2711.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. COLUMBIA;
RA Federici N.A., Palm C.J., Conway A.B., Kurtz D.B., Conway A.R.,
RA Au M., Araujo R., Buehler E., Dewar K., Feng J., Kim C., Li Y.,
RA Oji O., Osborne B.I., Shim P., Sun H., Toriumi M., Vysotskaya V.S.,
RA Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL: AC004122; AAC34329.1; -
DR InterPro: IPR002422; AA/rel. pmasee2.
DR Pfam: PF01490; Aa.trans. 1.
SQ SEQUENCE 475 AA; 51815 MW; B0DEB571E2FE7F92 CRC64;
Query Match 78.4%; Score 1974.5; DB 10; Length 475;
Best Local Similarity 75.4%; Pred. No. 6.4e-146;
Matches 364; Conservative 58; Mismatches 52; Indels 9; Gaps 2;
QY 1 MMSPTBEGNHSTABSGDAVTVSDPTKNDVEDERKRTGTMTLTAHITTAIVGSGVLSL 60
DB 1 MDAYN---NPSAVESGDA----AVKSYVDDBGKREKRTGTFWTAHITTAIVGSGVLSL 51
QY 61 AWAIAQLGWIAGTSTILIPSFITFTSTMLADCYRAPDPVTKRNTYMDVVRASYLGGRK 120
DB 52 AWAIAQLGWIAGTSTILIPSFITFTSTMLADCYRAPDPVTKRNTYMDVVRASYLGGRK 111
QY 121 VOLCGVAQGNLIGTVGTTITASISLVAVGKNCPHDKGHTADCTISNYPYMAVEGIIQ 180
DB 112 VOLCGVAQGNLIGTVGTTITASISLVAVGKNCYHDKGKAKGVSVPYMAAFGIQ 171
QY 181 VILSOLPNHKSFLSISMAVMSFTATATIGIALITVAGGKGTSMGTAVGDUVTAA 240
DB 172 IILSOLPNHKSFLSISMAVMSFTATATIGIALITVAGGKGTSMGTAVGDUVTAA 231
QY 241 OKIMPSFOAVGDIAPAYAVATVLIBIOTLRSSPAENKMKRASLVGSGTTFEYILCGC 300
DB 232 BKVMKLFQALIGDIARSYATTTILIEIOTLRSSPAENKMKRASLVGSGTTFEYILCGC 291
QY 301 IGYAAGNNAPOGFLTDFGFEFPMILIDPANACIAVHLIGAYOVFAOPIFQFVEKKNEN 360
DB 292 IGYAAGNNAPOGFLTDFGFEFPMILIDPANACIALHLIGAYOVFAOPFQFVEKKNEN 351
QY 361 YPDNKFITSESVNVPFLCKENISLFLVWRTAYVITTVVAMIPFPFNALIGLGAASF 420
DB 352 WQSNININKEYSKVPDLCKCRVNLFRLVWRTCYVVLITFVAMIPEFPFNALIGLGAFAF 411
QY 421 WPLTYVPFVEMHIAOTKIKKYSARWIALKTMCVCLIVSLAAGSIAGLISSVTKYKPF 480

DB 412 WPLTYVPFVEMHIAOTKIKKYSRRWIALNLVCLIVSALAAVGSIIIGLINSVSKYKPF 471
QY 461 RTM 483
DB 472 KML 474
RESULT 4
ID P92934 PRELIMINARY; PRT: 481 AA.
AC P92934
DT 01-MAY-1997 (TREMBlrel. 03, Created)
DT 01-MAY-1997 (TREMBlrel. 03, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Amino acid permease 6.
GN AAB6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. LANDSBERG; TISSUE=LEAF;
RX MEDLINE=96373203; PubMed=8776904;
RA Rentsch D., Hirner B., Schmelzer B., Frommer W.B.;
RT "Salt stress-induced proline transporters and salt stress-repressed
RT broad specificity amino acid permeases identified by suppression of a
RT yeast amino acid permease-targeting mutant.";
RL Plant Cell 8:1437-1446(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty P1 and TAC
RT clones.";
RL DNA Res. 7:31-63(2000).
DR EMBL: X95736; CAA65051.1; -
DR EMBL: AB025627; BAA97227.1; -
DR InterPro: IPR002422; AA/rel. pmasee2.
DR InterPro: IPR001395; Aldo/kel_red.
DR Pfam: PF01490; Aa.trans. 1.
DR PROSITE: PS00063; ALDOKE TO REDUCTASE 3; UNKNOWN 1.
SQ SEQUENCE 481 AA; 53020 MW; DEB1CA7348760AD1_CRC64;
Query Match 74.3%; Score 1871; DB 10; Length 481;
Best Local Similarity 73.6%; Pred. No. 7.8e-138;
Matches 345; Conservative 60; Mismatches 62; Indels 2; Gaps 2;
QY 15 ESGDAVTVSDPTKNDVEDERKRTGTMTLTAHITTAIVGSGVLSLAWAIAQLGWIAGTST 74
DB 11 QSPFHEIDGTKNKDEBDRDKRTGTMTGSAHITTAIVGSGVLSLAWAIAQLGWIAGTST 70
QY 75 IILSPTFTSTMLADCYRAPDPVTKRNTYMDVVRASYLGGRVQLCGVAQGNLIG 134
DB 71 VLMAFSFITFTSTMLADCYRAPDPVTKRNTYMEVVRASYLGGRVQLCGVAQGNLIG 130
QY 135 VVVGTTITASISLVAVGKNCPHDKGHTADCTISNYPYMAVGGIIQVLSOLPNHKSFL 194
DB 131 ITTGYTTTASISVAVKRCNCFHKGHNKCACTSNPFMIFAILIOTLISOLPNHLSM 190
QY 195 LSIMAAMVMSFTATATIGIALITVA-GGKVGKTSMTGTAVGDUVTAAQKIMPSFOAVGDI 253
DB 191 LSLIAAMVMSFTATATIGIALITVA-GGKVGKTSMTGTAVGDUVTAAQKIMPSFOAVGDI 250
QY 254 APAYAVATVLIBIOTLRSS-PAENKMKRASLVGSGTTFEYILCGTGYAAGNNAPOG 312
DB 251 APAYAVATVLIBIOTLRSS-PAENKMKRASLVGSGTTFEYILCGTGYAAGNNAPOG 310

```
QY 313 DFLTDFGFPFWLIDFANACIAVHLIGAYQVFAQPIQFQVEKKCNKNYPDNKFTITSEYS 372
Db 311 NFUTGFGFEPFWLIDFANVCIAVHLIGAYQVFCQIFQFVESQSAKRWPNKFTIGEYK 370
QY 373 VNVPLGKFNISFLRWRTAYVVITTVAMIPFFFNAILGLLGAASFPLTVYFFVEMH 432
Db 371 IHVPCGDFSFNLRVLWRTSYVVVTVAVAMIPFFFNDFLGLLGAASFPLTVYFFVEMH 430
QY 433 IAQTKIKKYSARWIAKTMCYCLIVSLLLAAGSIAGLISSVYTYXPR 481
Db 431 IAQKIPKFSFTWLKILSWTCTFIVSLVAAGSVQGLIQSLKDFPFQ 479

RESULT 5
ID P93561 PRELIMINARY; PRT; 385 AA.
AC P93561;
DT 01-MAY-1997 (TrEMBLrel. 03, Created)
DT 01-MAY-1997 (TrEMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Amino acid transporter (Fragment).
GN AAP1.
OS Solanum tuberosum (Potato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; euasterids I; Solanales; Solanaceae; Solanum.
OX NCBI_TaxID=4113;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. DESIREE; TISSUE=TUBER;
RA Kwart M., Laubner M., Heinike D., Rentsch D., Frommer W.B.;
RT "An increased C/N ratio in potato tubers due to antisense inhibition
of the leaf H+/amino acid transporter StaAP1.";
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y09825; CAA70968.1; -
DR InterPro; IPR002422; AA/rel_prmease2.
DR Pfam; PF01490; Aa_trans; 1.
FT NON_TER 1
FT NON_TER 385
SQ SEQUENCE 385 AA; 41942 MW; 5EF4B87FAD7F6A9 CRC64;

Query Match 63.1%; Score 1589.5; DB 10; Length 385;
Best Local Similarity 74.5%; Pred. No. 5.5e-116;
Matches 287; Conservative 55; Mismatches 42; Indels 1; Gaps 1;

QY 47 HIITAVIGSGVLSLAWAIAQLGWIAGTSTLLIFSPITYFTSTMLADCYRAPDPVTKRNY 106
Db 1 HIITAVIGSGVLSLAWAIAQLGWVAGPAVLPAFSFITYFTSTLLADCYRSPGPGISKRNY 60
QY 107 TYMDVRSYLGRKVLGCVQVAGYGNLIGVTGYTTTASISLVAVGKSNCFHDKGHTADCT 166
Db 61 TYMDVRSYHLGGVYVLCGIAQYANLVGTGYTTTASISLVAVKRSNCFHKNHGHEASCS 120
QY 167 ISNYPYMAVFGIIVQLISQIPNPKLSFLSILAAVMSFTYATIGLAIATVAG-OKVGK 225
Db 121 IESYPMIIFAVLIQVLSQIPNPKLSWLSILAAVMSFTYASIGLISLAKASGVGHVK 180
QY 226 TSMGTGAVGVDTAAQKWRSFQVAGDIAFAYAYATVLIIEIQTLLRSSPAENKAMKRASL 285
Db 181 TALTVGVGVGVDSGTEKWRSFQVAGDIAFAYAYSTVLIIEIQTLLKSPSPSEKVMKRASL 240
QY 286 VGVSTTTFFYLCCGIGYAAFGNAPGDELTDGFPFPEWLDIDFANACIAVHLIGAYQVF 345
Db 241 AGVSTTTFLYVLCGTIGYAAFGNAPGDELTDGFPFPEWLDIDFANVCIAVHLVAGAYVF 300
QY 346 AQPIQFQVEKKCNKNYPDNKFTITSEYSVNVPLGKFNISFLRWRTAYVVITTVAMIP 405
Db 301 CQPLYGFEVGRSERPDSKFTITSEYAMQVPCGTTNLNLFRLVWRTTIVTAVIAMIP 360
QY 406 PFFNAILGLIGAASFPLTVYFFVE 430
Db 361 PFFNDFLGLIGAASFPLTVYFFVE 385
```

```
RESULT 6
ID Q82044 PRELIMINARY; PRT; 486 AA.
AC Q82044;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Amino acid carrier.
GN AAP1.
OS Ricinus communis (Castor bean).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids I; Malpighiales; Euphorbiaceae; Ricinus.
OX NCBI_TaxID=3988;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV. SANGUINEOUS;
RX MEDLINE=98405925; PubMed=9733991;
RA Marvier A.C., Neelam A., Bick J.A., Hall J.L., Williams L.E.;
RT "Cloning of a cDNA coding for an amino acid carrier from Ricinus
communis (RcAAP1) by functional complementation in yeast: kinetic
analysis, inhibitor sensitivity and substrate specificity.";
RL Biochim. Biophys. Acta 1373:321-331(1998).
DR EMBL; AJ007574; CAA07563.1; -
DR InterPro; IPR002422; AA/rel_prmease2.
DR Pfam; PF01490; Aa_trans; 1.
SQ SEQUENCE 486 AA; 53113 MW; 882E0ADA50D18F16 CRC64;

Query Match 61.2%; Score 1540.5; DB 10; Length 486;
Best Local Similarity 58.2%; Pred. No. 4.8e-112;
Matches 285; Conservative 82; Mismatches 106; Indels 17; Gaps 2;

QY 5 NTEGHNHSTAESGDAYTVSDPTKN-----VDEGREKRTGTWLTASAHITAVIG 54
Db 4 NTAANKH-----PHQVFDVSNMQTVGSKWLDGDDGRTKRTGVWTASAHITAVIG 56
QY 55 SGVLSLAWAIAQLGWIAGTSTLLIFSPITYFTSTMLADCYRAPDPVTKRNYTYMDVRS 114
Db 57 SGVLSLAWAIAQLGWIAGPAVMFLFSLVITYTSTLLSACYRSGDPVNGKRNYYTMDAVRT 116
QY 115 YLGRKVLGCVQVAGYGNLIGVTGYTTTASISLVAVGKSNCFHDKGHTADCTISNYPYMA 174
Db 117 NLGGAKVLCGFVQVYLNLFQVAGIYTIASSISWMAIKRSNCFHKSQGNPCHINAPYMI 176
QY 175 VEGIIVQLISQIPNPKLSFLSILAAVMSFTYATIGLAIATVAGGKVGKTSMTGTAVG 234
Db 177 AFGIAEIIFSQIPDFDLQWLWLSILAAVMSFTYTGILGLGIAQVVENKAMGSVTGISIG 236
QY 235 VDVTAQAQKWRSFQVAGDIAFAYAYATVLIIEIQTLLRSSPAENKAMKRASLVGVSTTTFF 294
Db 237 ANVTPTQIKWRSFQALGDIAFAYSIIILIEIQTIVRSPSPSESKTKKATLISVAVTTLF 296
QY 295 YLCCGIGYAAFGNAPGDELTDGFPFPEWLDIDFANACIAVHLIGAYQVFAQPIQFVE 354
Db 297 YMLCCGFGYAAFGDMSPNLITGFGFYNPYWLDDIANVAIVHLVAGAYQVTCQPLFAFVE 356
QY 355 KCCNRYNDKNFPTITSEYSVNVPLGKFNISFLRWRTAYVVITTVAMIPFFFNAILGL 414
Db 357 KAAQRYPDSGFITKDIKIPVGPFRPNLFRSVWRTLFVVVFTTVISMLLPFFNDIVGL 416
QY 415 TGAASFPLTVYFFVEMHIAQTKIKKYSARWIAKTMCYCLIVSLLLAAGSIAGLISV 474
Db 417 LGALEWFLTVYFFVEMYIAQKIPKFSFTWLKILSWTCTFIVSLVAAGSVQGLIAGVGD 476
QY 475 KTYKDFPRTMH 484
Db 477 KSVKEPQISY 486

RESULT 7
ID Q39134 PRELIMINARY; PRT; 476 AA.
Q39134
```

AC Q39134;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-JAN-1999 (TRENBLREL. 09, Last sequence update)
 DT 01-MAR-2002 (TRENBLREL. 20, Last annotation update)
 DE Amino acid permease (Amino acid carrier, putative).
 OS AAP3 OR F2P24.9.
 GN Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fischer W.N.;
 RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=95332342; PubMed=7608199;
 RA Fischer W.N., Kwart M., Hummel S., Frommer W.B.;
 RT "Substrate specificity and expression profile of amino acid
 RT transporters (AAPs) in Arabidopsis."
 RL J. Biol. Chem. 270:16315-16320(1995).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Fischer W.N.;
 RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA STRAIN=CV. COLUMBIA;
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,
 RA Maiti R., Ronning C.M., Koo H., Fujii C.Y., Utecherback T.R.,
 RA Barnstead M.E., Bowman C.L., White O., Niernan W.C., Fraser C.M.;
 RT "Arabidopsis thaliana chromosome 1 BAC P2P24 genomic sequence."
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X77499; CA545630.1; -
 DR EMBL; AC078898; AAG29203.1; -
 DR InterPro; IPR002422; AA/re1_pmease2.
 DR Pfam; PF01490; Aa_trans; 1.
 SQ SEQUENCE 476 AA; 52037 MW; FEC8C42037804F67 CRC64;
 Query Match 59.7%; Score 1504; DB 10; Length 476;
 Best Local Similarity 57.6%; Pred. No. 3.3e-109;
 Matches 273; Conservative 97; Mismatches 100; Indels 4; Gaps 4;
 QY 10 NSTAESGDAYVYSDPTKVNDEDEGRKRGTMLTASAHITNTGSGVSLAMAIQOLG 69
 DB 4 NQOTVLAVD-MPQTGSKYLDGDKRKRKRGSVTASAHITNTAVTSGSVLSLAATAQOLG 62
 QY 70 IAGTSLILFSFTYFTSTMLADCYRAPDPVTGKRNVTYMDVRSYIGRKVOLCGVAQY 129
 DB 63 LAGPVMMLFSANTYFTSSSLAACYRSGDPISGKRNVTYMDVRSNLGKVTLCGIVQY 122
 QY 130 GNLIGVTVGTTTASISLVAAGKSNCFPHDKGHTADCTISNYPYMAVFGIIOVILSQIPNF 189
 DB 123 LNIFFGVAIGVTTASAISSMAIKRSCFHKSGKDPCHMNSNPMIAGVLQIILFSQIPDF 182
 QY 190 HKLSFSLMAAVSFTYATIGTGLATA-TYAGKRGKTSMTGTAVGVDTAAKIRSRQ 248
 DB 183 DQIMMISILAAVNSFTYSSAGLGLIAQVAVNKKV-KGSLTGISIA-VTEGKIKIRTRQ 240
 QY 249 AVGDIAFAVAATVLLIEIDTLRSSPAENKAMRASLVGSTTTFYILGCGICGVAFGN 308
 DB 241 ALGDIAFAVSYSILILEIDTIVKSPSEKTKKATLVSVTMTMYMLCGCMGYAALFGD 300
 QY 309 NAFGDFLTPGFEFPWLIDFANACIAVHLIGAYQVPAQDIFQFVEKKCRNRYPNKKFT 368
 DB 301 LSPGNLITLGGFYNPYMLDIAANAIVHLIGAYQVCCPLFAFIRKQSIGPDSSEFLA 360
 QY 369 SESVAVPFLGKKNISLFRVWRTAVVTTVYAMIFPPFNALIGLIGASFPPLTVYFP 428
 DB 361 KDKITLPFGKPLRLNVFRIMRTVEVITTVISMLLPFNDVAVGALGAFPLTVYFP 420
 QY 429 VEMHIAQTKIKYSARWIALKTKCYVCLIVSLAAAGSISGLISSVTKYRPFRT 482

DB 421 VEMYIAOKKIPRWSRWCLQVFSIGCLVSSIAAAGSIAGVLIDLKSKYKPPRS 474
 RESULT 8
 ID Q38967 PRELIMINARY; PRT; 493 AA.
 AC Q38967;
 DT 01-NOV-1996 (TRENBLREL. 01, Created)
 DT 01-NOV-1996 (TRENBLREL. 01, Last sequence update)
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
 DE Amino acid permease (Amino acid transport protein AAP2)
 DE (AUS909220/T5E8_20).
 GN AAP2 OR T5E8.20.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=94108493; PubMed=8281191;
 RA Kwart M., Hinner B., Hummel S., Frommer W.B.;
 RT "Differential expression of two related amino acid transporters with
 RT differing substrate specificity in Arabidopsis thaliana."
 RL Plant J. 4:993-1002(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Murphy G., Ridley P., Hudson S., Bancroft I., Mewes H.W.,
 RA Rudd S., Lemcke K., Mayer K.F.X.;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Cheuk R., Chen H., Kim C.J., Meyers M.C., Shim P., Banh J.,
 RA Bowser L., Carninci P., Chang E., Dale J.M., Goldsmith A.D.,
 RA Hayashizaki Y., Ishida J., Jones T., Kanuya A., Katlin-Neumann G.,
 RA Kawai U., Lam B., Lee J.M., Lin J., Miranda M., Natusaka M.,
 RA Nguyen M., Onodera C.S., Palm C.J., Quach H.L., Sakurai T., Satou M.,
 RA Seki M., Southwick A., Tang C.C., Toriumi M., Wu H.C., Yamada K.,
 RA Yamamura Y., Yu G., Yu S., Shinzaki K., Davis R.W., Theologis A.,
 RA Ecker J.R.;
 RT "Arabidopsis cDNA clones";
 RT Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; X71787; CAA50672.1; -
 DR EMBL; AL391712; CAC05448.1; -
 DR EMBL; AY090341; AAL91247.1; -
 DR InterPro; IPR002422; AA/re1_pmease2.
 DR Pfam; PF01490; Aa_trans; 1.
 SQ SEQUENCE 493 AA; 54147 MW; 7CA882E3E011C5C3 CRC64;
 Query Match 59.6%; Score 1501; DB 10; Length 493;
 Best Local Similarity 56.7%; Pred. No. 5.9e-109;
 Matches 271; Conservative 94; Mismatches 111; Indels 2; Gaps 2;
 QY 8 GHNHSTAESGDAYVSDPTKVNDEDEGRKRGTMLTASAHITNTAVTSGSVLSLAIAQOL 67
 DB 17 GHQVFDVASHDFPQPAKCFDDGRLKRTGTWMTASAHITNTAVTSGSVLSLAIAQOL 76
 QY 68 GNIAGTSLILFSFTYFTSTMLADCYRAPDPVTGKRNVTYMDVRSYIGRKVOLCGVA 127
 DB 77 GNIAGPVMMLFSANTYFTSSSLAACYRSGDPISGKRNVTYMDVRSNLGKVTLCGIVQY 126
 QY 128 QYGNLIGVTVGTTTASISLVAAGKSNCFPHDKGHTADCTISNYPYMAVFGIIOVILSQIP 187
 DB 137 QYTNLFGIAGTITTAISMAIKRSCFHKSGKDPCHMNSNPMIAGVLQIILFSQIPDF 196
 QY 188 NPHKLSFSLMAAVSFTYATIGTGLATA-TYAGKRGKTSMTGTAVGVDTAAKIRSRQ 247
 DB 197 DFDQIMWISIVAAVNSFTYSSAGLGLIAQVAVNKKV-KGSLTGISIA-VTEGKIKIRTRF 255

```
QY 248 QAVGDIAPAYATVLEIOTDLRSPAENKAMKRAVLGVSTTTTFFYLGCIGYAARG 307
Db 256 QALGDIAPAYSIVLVLEIOTDLRSPAENKAMKRAVLGVSTTTTFFYLGCIGYAARG 315
QY 308 NNAPGDLTDFGFEFFWLTIDFANACIAVHLIGAYQVFAQPIQFQVEKKNRNYPDKFI 367
Db 316 DAAPGNLLTCGFGFYNFWLLDIAAAVHLVGVAYQVFAQPIFAFIEKSAERYPNDFL 375
QY 368 TSYSVNVVP-FLGKNISLRLVWRTAYVVTIVVAMIPFFFNAILGLIGAASFWLTVY 426
Db 376 SKEFEIRIPGKSPYKVNFRMYRSGFVVTTTIVSMLMPFFNDVVGILGALGFNLT 435
QY 427 FVEMHIAOTKIYKSGARMTALCMVCVLIVSLLAAGSIAGLISSVKYKPPRTWH 484
Db 436 FVEMYIKQKVKESWTRVCLQWLSVACLIVSVVAGVSGIAGVMDLKVYKPKFTY 493

RESULT 9
Q8RWAB PRELIMINARY; PRT; 466 AA.
AC Q8RWAB;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE Amino acid transporter AAP4.
GN ATSG63850.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayaishizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shimizu P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.
RA Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RL EMBL; AY093224; AM13224; -.
SQ SEQUENCE 466 AA; 51359 MW; 11B1D5329C630D89 CRC64;

Query Match 59.4%; Score 1496; DB 10; Length 466;
Best Local Similarity 58.0%; Pred. No. 1.4e-108;
Matches 266; Conservative 99; Mismatches 92; Indels 2; Gaps 2;

QY 27 KNVEDGREGKRTGTWLTASAHITAVIGSGVLSLAWAIAQLGWIAGTSTILLIPSPITYFT 86
Db 9 KCFDDGRLKRSCTVWTASAHITAVIGSGVLSLAWAIAQLGWIAGTPTVLLFSFVTVYS 68

QY 87 STMLADCYRAPDPVTGKRNVTYMDVRSYLGGRKVLGCVQAGNIGLVGTGYTTTASIS 146
Db 69 STLLSDCYRTGDPVSGKRNVTYMDVRSYLGGRKVLGCVQAGNIGLVGTGYTTTASIS 128

QY 147 LVAVGKSNCFHDKGHTADCTISNYPMAVFGIIVLSQIPNFHKLSPFLSIWAAVMSFTY 206
Db 129 MMAIKRSNCFHESGGKNPCHMSSNPYIMFGVTEILLQIKDFQIWWLSIVAAIMSFY 188

QY 207 ATIGIGLATVAGGKGTMTGTAVGVDDVTAQKIRSFQAVGDIAPAYATVLEI 266
Db 189 SAIGLALGIIQVAANGVVGKSLTGISGA-VTQTKIWRFTQALGDIAPAYSIVVLEI 247

QY 267 QOTLRSSPAENKAMKRAVLGVSTTTTFFYLGCIGYAAGFNAPGDLTDFGFEFFWL 326
Db 248 QOTVRSPPAESKTMKTIATRIASIAVTTTFYMLCGCMGYAAGDKAPGNLLTGFGFYNPFWL 307

QY 327 IDPANACIAVHLIGAYQVFAQPIQFQVEKKNRNYPDKFIQSEYVNV-PLGKNISL 385
Db 308 LDVANAIAVHLVGVAYQVFAQPIFAFIEKQAAAPFSDLVTKETIRIPGRSPYKVN 367

QY 386 FRLVWRTAYVVTIVVAMIPFFFNAILGLIGAASFWLTVYFPVEMHIAQTKIKKYSARW 445
Db 368 FRAVYRSFVVLTVISMLMPFFNDVVGILGALGFNLTGFGFYNPFWL 427

QY 446 IALKTMCVCLIVSLLAAGSIAGLISSVKYKPPRTWH 484
Db 428 VCLQWLSGCLMITLVAGVSGIAGVMDLKVYKPKFTY 466
```

```
Db 368 FRAVYRSFVVLTVISMLMPFFNDVVGILGALGFNLTGFGFYNPFWL 427
QY 446 IALKTMCVCLIVSLLAAGSIAGLISSVKYKPPRTWH 484
Db 428 VCLQWLSGCLMITLVAGVSGIAGVMDLKVYKPKFTY 466

RESULT 10
Q9FN04 PRELIMINARY; PRT; 466 AA.
AC Q9FN04;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE Amino acid transporter AAP4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura Y., Sato S., Kaneko T., Kotani H., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. III.
RT Sequence features of the regions of 1,191,918 bp covered by seventeen
RT physically assigned P1 clones.";
RL DNA Res. 4:401-414 (1997).
DR EMBL; AB007646; BAB11033.1; -.
DR InterPro; IPR004422; AA/rel_pmasease2.
DR Pfam; PF01490; Aa_trans; 1.
SQ SEQUENCE 466 AA; 51428 MW; 11B1D5332C63BD88 CRC64;

Query Match 59.3%; Score 1495; DB 10; Length 466;
Best Local Similarity 58.0%; Pred. No. 1.6e-108;
Matches 266; Conservative 98; Mismatches 93; Indels 2; Gaps 2;

QY 27 KNVEDGREGKRTGTWLTASAHITAVIGSGVLSLAWAIAQLGWIAGTSTILLIPSPITYFT 86
Db 9 KCFDDGRLKRSCTVWTASAHITAVIGSGVLSLAWAIAQLGWIAGTPTVLLFSFVTVYS 68

QY 87 STMLADCYRAPDPVTGKRNVTYMDVRSYLGGRKVLGCVQAGNIGLVGTGYTTTASIS 146
Db 69 STLLSDCYRTGDPVSGKRNVTYMDVRSYLGGRKVLGCVQAGNIGLVGTGYTTTASIS 128

QY 147 LVAVGKSNCFHDKGHTADCTISNYPMAVFGIIVLSQIPNFHKLSPFLSIWAAVMSFTY 206
Db 129 MMAIKRSNCFHESGGKNPCHMSSNPYIMFGVTEILLQIKDFQIWWLSIVAAIMSFY 188

QY 207 ATIGIGLATVAGGKGTMTGTAVGVDDVTAQKIRSFQAVGDIAPAYATVLEI 266
Db 189 SAIGLALGIIQVAANGVVGKSLTGISGA-VTQTKIWRFTQALGDIAPAYSIVVLEI 247

QY 267 QOTLRSSPAENKAMKRAVLGVSTTTTFFYLGCIGYAAGFNAPGDLTDFGFEFFWL 326
Db 248 QOTVRSPPAESKTMKTIATRIASIAVTTTFYMLCGCMGYAAGDKAPGNLLTGFGFYNPFWL 307

QY 327 IDPANACIAVHLIGAYQVFAQPIQFQVEKKNRNYPDKFIQSEYVNV-PLGKNISL 385
Db 308 LDVANAIAVHLVGVAYQVFAQPIFAFIEKQAAAPFSDLVTKETIRIPGRSPYKVN 367

QY 386 FRLVWRTAYVVTIVVAMIPFFFNAILGLIGAASFWLTVYFPVEMHIAQTKIKKYSARW 445
Db 368 FRAVYRSFVVLTVISMLMPFFNDVVGILGALGFNLTGFGFYNPFWL 427

QY 446 IALKTMCVCLIVSLLAAGSIAGLISSVKYKPPRTWH 484
Db 428 VCLQWLSGCLMITLVAGVSGIAGVMDLKVYKPKFTY 466
```

```

RESULT 11
ID Q39135 PRELIMINARY; PRT; 466 AA.
AC Q39135;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Amino acid transporter.
GN AAP4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustrosia 11; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Fischer W.N.;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA MEDLINE=9532342; PubMed=7608199;
RA Fischer W.N., Kwart M., Hummel S., Frommer W.B.;
RT "Substrate specificity and expression profile of amino acid
RT transporters (AAPs) in Arabidopsis."
RT J. Biol. Chem. 270:16315-16320(1995).
DR EMBL; X77500; CAA54631.1; -.
DR InterPro; IPR002422; AA/rel_pmease2.
DR Pfam; PF01490; Aa_trans; 1.
SQ SEQUENCE 466 AA; 51470 MW; 1538DF32C63B778 CRC64;

Query Match 59.3%; Score 1494; DB 10; Length 466;
Best Local Similarity 58.0%; Pred. No. 1.9e-108;
Matches 266; Conservative 98; Mismatches 93; Indels 2; Gaps 2;

QY 27 KNVDEGRKRGKRTGLTAHITTAIVGSGVSLAAIAIQLGHIAGTSLILFSFTTYT 86
DB 9 KCFDDGRLKRGSGVTASAHITTAIVGSGVSLAAIAIQLGHIAGTSLILFSFTTYT 68
QY 87 STMADCYAPPVTKRYTMDVRSYLGGRKVLGCAVOYGNLIGTVGTTASIS 146
DB 69 STLSDCTRTGPPVSGKRYTMDVRSYLGGRKVLGCAVOYGNLIGTVGTTASIS 128
QY 147 LVAVGKSNCFHDKGTADCTISNYPYAVFGIIOVLISQIPNFHKLSPISMAVMSFTY 206
DB 129 WMAIKRSNCFHSGKGNPCMHMSNPFYIMFGVTEILSLQIKDPDQIWMISYVAIMSFY 188
QY 207 ATITGLAIATYAGKVGKTSMTGTAAGVDTAAQKIMRSFOAVGDIAPAYATVLIIEI 266
DB 189 SAIGLALGIIQVAANGVVGSLTGISIGA-VIQTKIMRTFQALGDIAPAYSVVLIEI 247
QY 267 OPTLSSPAENKAMKRASLVGSTTFFYLICGCGIYAAFGNNAKPDPLTDGFFPFWL 326
DB 248 QDTVSPSPAESKTMKATISIAVTTFFYMLCCGCKVYAAFGDAKPNLLTGFEFYPFWL 307
QY 327 IDPANACIAVHLIGAYQVAPQIFQVEKKCNKNRNPDKFTSEYVNP-FLGKNISL 385
DB 308 LVAANAIAIYHIVGAAQVAPQIFAFIEKQIARFPDSDLVTEYERIRIPGFSPKAVV 367
QY 386 FFLWRTAVVITTVVAMIFPPFNALIGLGAASFWPLTVYFVPEVEMHIAQTKIKYSARW 445
DB 368 FFAVYVSGVIVLTIVSMIPFENDVIGLIGFWPLTVYFVPEVEMHIAQTKIKYSARW 427
QY 446 IALKTMCVCLIVSLAAGSIAGLSSVKTKEPRTMH 484
DB 428 VCIQMLSCGCLMTITLVAGVSIAGVMDLKVKKPKFTTY 466

RESULT 12
ID Q9ZPM7 PRELIMINARY; PRT; 379 AA.
AC Q9ZPM7;
DT 01-MAY-1999 (TREMBlrel. 10, Created)
DT 01-MAY-1999 (TREMBlrel. 10, Last sequence update)

```

```

DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE Amino acid transporter (Fragment).
GN AAP1.
OS Nepenthes alata (Winged pitcher plant).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Caryophyllidae; Caryophyllales; Nepentaceae; Nepenthes.
OX NCBI_TaxID=4376;
RN [1]
RP SEQUENCE FROM N.A.
RA Tissue=Pitcher;
RC Schulze W., Frommer W.B., Ward J.M.;
RT "Transporters for ammonium, amino acids and peptides are expressed in
RT pitchers of the carnivorous plant Nepenthes."
RT Plant J. 17:101-110(1999).
DR EMBL; AF080542; AADI6013.1; -.
DR InterPro; IPR002422; AA/rel_pmease2.
DR Pfam; PF01490; Aa_trans; 1.
FT NON_TER 1
FT NON_TER 379
SQ SEQUENCE 379 AA; 41553 MW; FF2452DFB4B0CFDA CRC64;

Query Match 58.9%; Score 1484.5; DB 10; Length 379;
Best Local Similarity 71.8%; Pred. No. 8.3e-108;
Matches 272; Conservative 48; Mismatches 58; Indels 1; Gaps 1;

QY 51 AVYGGVSLAAIAIQLGHIAGTSLILFSFTTYSTMADCYAPPVTKRYTMD 110
DB 1 AVYGGVSLAAIAIQLGHIAGTSLILFSFTTYSTMADCYAPPVTKRYTMD 60
QY 111 VYRSLGGRKVLGCAVOYGNLIGTVGTTASISLVAVGKSNCFHDKGTADCTISNY 170
DB 61 VYKAVILGGRKVLGCAVOYGNLIGTVGTTASISLVAVGKSNCFHDKGTADCTISNY 120
QY 171 PMAVFGIIOVLISQIPNFHKLSPIS-IMAAVMSFTYATITGLAIATYAGKVGKTSMT 229
DB 121 PMAVFGIIOVLISQIPNFHKLSPIS-IMAAVMSFTYATITGLAIATYAGKVGKTSMT 180
QY 230 GTAVGVDVTAQKIMRSFOAVGDIAPAYATVLIIEIOTLSSPAENKAMKRASLVGS 289
DB 181 GTAVGVDVTAQKIMRSFOAVGDIAPAYATVLIIEIOTLSSPAENKAMKRASLVGS 240
QY 290 TTFYFYLICGCGIYAAFGNNAKPDPLTDGFFPFWLIDPANACIAVHLIGAYQVAPQI 349
DB 241 PPHSFYVLCGCGIYAAFGNNAKPDPLTDGFFPFWLIDPANACIAVHLIGAYQVAPQI 300
QY 350 FQVEKKCNKNRNPDKFTSEYVNPFLGKNISLFLWRTAVVITTVVAMIFPPFN 409
DB 301 FAFVSGVCRDMPENKFTREHPIEVPEVGVYVNLFLWRTAVVITTVVAMIFPPFN 360
QY 410 ATITGLAIATYAGKVGKTSMTGTAAGVDTAAQKIMRSFOAVGDIAPAYATVLIIEI 428
DB 361 DFLGLIGASFWPLTVYFP 379

RESULT 13
ID Q8RZP7 PRELIMINARY; PRT; 488 AA.
AC Q8RZP7;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
DE Putative amino acid transport protein.
GN B106SE10.17.
OS Oryza sativa (Japanese cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. NIPONBARE;
RA Sasaki T., Matsunoto T., Yamamoto K.;

```


PS Claim 5; SEQ ID NO 596; 261pp + Sequence Listing; English.

XX The invention relates to identifying target proteins
CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

SQ Sequence 485 AA;

Query Match 100.0%; Score 2519; DB 23; Length 485;
Best Local Similarity 100.0%; Pred. No. 5.9e-266;
Matches 485; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 M K S F N T E G H N H S T A E S G D A Y T V S D P T K N V D E D G R E K R T G T W L T A S A H I I T A V I G S G V L S L 60

Db 1 M K S F N T E G H N H S T A E S G D A Y T V S D P T K N V D E D G R E K R T G T W L T A S A H I I T A V I G S G V L S L 60

Qy 61 A W A I A Q L G W I A G T S I L L I F S F I T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G R K 120

Db 61 A W A I A Q L G W I A G T S I L L I F S F I T Y T F T S T M L A D C Y R A P D P V T G K R N Y T Y M D V V R S Y L G R K 120

Qy 121 V O L C G V A Q Y G N L I G V T V G Y T T I T A S I S L V A G K S N C F H D K G H T A D C T I S N Y P M A V F G I I Q 180

Db 121 V O L C G V A Q Y G N L I G V T V G Y T T I T A S I S L V A G K S N C F H D K G H T A D C T I S N Y P M A V F G I I Q 180

Qy 181 V I L S Q I P N F H K L S F L S I M A A V M S F T A T I G I G L A I A T V A G G K V G K T S M T G T A V G D V T A A 240

Db 181 V I L S Q I P N F H K L S F L S I M A A V M S F T A T I G I G L A I A T V A G G K V G K T S M T G T A V G D V T A A 240

Qy 241 Q K I W R S F Q A V G D I A F A Y A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300

Db 241 Q K I W R S F Q A V G D I A F A Y A T V L I E I Q D T L R S S P A E N K A M K R A S L V G V S T T T F F Y I L C G C 300

Qy 301 I G Y A A F G N N A P G D F L T D F G F F P F W L I D F A N A C I A V H L I G A Y Q V F A Q P I F Q P V E K K C N R N 360

Db 301 I G Y A A F G N N A P G D F L T D F G F F P F W L I D F A N A C I A V H L I G A Y Q V F A Q P I F Q P V E K K C N R N 360

Qy 361 Y P D N K F I T S E Y S V N V P F L G K N I S L F R L V R T A Y V V I T T V A M I P F F F N A I L G L I G A A S F 420

Db 361 Y P D N K F I T S E Y S V N V P F L G K N I S L F R L V R T A Y V V I T T V A M I P F F F N A I L G L I G A A S F 420

Qy 421 W P L T V Y F P V E M H I A Q T K I K Y S A R W I A L K T M C V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480

Db 421 W P L T V Y F P V E M H I A Q T K I K Y S A R W I A L K T M C V C L I V S L L A A A G S I A G L I S S V K T Y K P F 480

Qy 481 R T M H E 485

Db 481 R T M H E 485

RESULT 2

AAAR47812

ID AAAR47812 standard; Protein; 485 AA.

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

PF 05-JUL-1992; 92DE-4222315.

XX

PR 05-JUL-1992; 92DE-4222315.

XX

PA (GENB-) INST GENBIOLOGISCHE FORSCHUNG.

XX

PI Frommer W;

XX

XX

DR WPI; 1994-017036/03.

XX

DR N-PSDB; AAQ55058.

XX

XX

PT DNA coding for aminoacid transporter proteins - used for

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX Arabidopsis thaliana protein fragment SEQ ID NO: 2950.
DE
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EP1033405-A2.
XX
XX
PD 06-SEP-2000.
XX
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0128845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148665.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.

PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147130.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160747.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.

PR 29-OCT-1999; 99US-0162142.

Query Match 81.6%; Score 2056; DB 21; Length 397;
Best Local Similarity 99.7%; Pred. No. 1.7e-215;
Matches 396; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 89 MLADCYRAPDPVTGKNTYMDVVRVYLGGRKVLQCGVAQYGNLIGVTVGYTTASISLV 148
DB 1 MLADCYRAPDPVTGKNTYMDVVRVYLGGRKVLQCGVAQYGNLIGVTVGYTTASISLV 60

QY 149 AVGKSNCFHDKGHTADCTISNYPYMAVFGIIQVLSQIPNFHKLSPLSINAAVMSFTYAT 208
DB 61 AVGKSNCFHDKGHTADCTISNYPYMAVFGIIQVLSQIPNFHKLSPLSINAAVMSFTYAT 120

QY 209 IGIGLAIAIVAGGKVGKTSMTGTAVGVDTAAQKIWRSTQAVGDIAPAFAYATVLEIQD 268
DB 121 IGIGLAIAIVAGGKVGKTSMTGTAVGVDTAAQKIWRSTQAVGDIAPAFAYATVLEIQD 180

QY 269 TLRSSPAENKAMKRAKSLVGVSTTTFFVYILCGCIGYAAFGNNAAGDFLTDGPRPEPFWLID 328
DB 181 TLRSSPAENKAMKRAKSLVGVSTTTFFVYILCGCIGYAAFGNNAAGDFLTDGPRPEPFWLID 240

QY 329 FANACIAVHLIGAYQVFAQPIFOFVEBKCNRNYPDNKFTITSEYVNVVPLGKFNISLFL 398
DB 241 FANACIAVHLIGAYQVFAQPIFOFVEBKCNRNYPDNKFTITSEYVNVVPLGKFNISLFL 300

QY 389 VVRTAVVVTVVVAMTFFPNAILGLIGAASFWPLTVYPPVEMHIAOTKIKKYSARWIAL 448
DB 301 VVRTAVVVTVVVAMTFFPNAILGLIGAASFWPLTVYPPVEMHIAOTKIKKYSARWIAL 360

QY 449 KTMCVYCLIVSLAAAGSIAGLISVVKTKPPRTMHE 485
DB 361 KTMCVYCLIVSLAAARSIAGLISSVVKTKPPRTMHE 397

RESULT 5
ABB90933
ID ABB90933 standard; Protein; 475 AA.
XX AC ABB90933;
XX
DT 31-MAY-2002 (first entry)
DE Herbicidally active polypeptide SEQ ID NO 144.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 144; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant

CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
XX useful as herbicides.

SQ Sequence 475 AA;

Query Match 78.4%; Score 1974.5; DB 23; Length 475;
Best Local Similarity 75.4%; Pred. No. 1.7e-208;
Matches 364; Conservative 58; Mismatches 52; Indels 9; Gaps 2;

QY 1 MGSFTEGHSHSTAESDAYTVSDTKNVDEGRKRTGTWLTASAHITAVIGSVLSL 60
DB 1 MDAYN----NPSAVEGDA-----AVKSVDDGGRKRTGTFTWTAHITAVIGSVLSL 51

QY 61 AWAIQLGWIAGTSTILLIFSFITYTSTMLADCYRAPDPVTGKNTYMDVVRVYLGGRK 120
DB 52 AWAIQLGHWAGTTLVAFALITYTSTLLADCYRSPDSITGTRNRYNMGVVRVYLGGRK 111

QY 121 VOLCGVAQYGNLIGVTVGYTTASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 180
DB 112 VOLCGVAQYGNLIGVTVGYTTASISLVAVGKSNCFHDKGHTADCTISNYPYMAVFGIIQ 171

QY 181 VILSQIPNFHKLSPLSINAAVMSFTYATIGTGLATATVAGGKVGKTSMTGTAVGVDTAA 240
DB 172 IILSQLPNFHLSPLSINAAVMSFTYATIGTGLATATVAGGKVGKTSMTGTAVGVDTAA 231

QY 241 QKIMRSFOAVGDIAPAFAYATVLEIQDTLRSSPAENKAMKRAKSLVGVSTTTFFVILCGC 300
DB 232 EKVKLQFAIGDIAFSYAFITLIEIQDTLRSSPAENKAMKRAKSLVGVSTTTFFVILCGC 291

QY 301 IGYAAGNNAPGDFLTDGPFEPFWLIDFANACIAVHLIGAYQVFAQPIFOFVEBKCNRN 360
DB 292 IGYAAGNNAPGDFLTDGPFEPFWLIDFANACIAVHLIGAYQVFAQPIFOFVEBKCNRN 351

QY 361 YPDNKFITSEYVNVVPLGKFNISLFLVWRTAVVVTVVVAMTFFPNAILGLIGAAS 420
DB 352 WPOSNPFINKSEYVNVVPLGKFNISLFLVWRTAVVVTVVVAMTFFPNAILGLIGAAS 411

QY 421 WPLTVYPPVEMHIAOTKIKKYSARWIALKTMCVYCLIVSLAAAGSIAGLISVVKTKPP 480
DB 412 WPLTVYPPVEMHIAOTKIKKYSARWIALKTMCVYCLIVSLAAAGSIAGLISVVKTKPP 471

QY 481 RTM 483
DB 472 KNL 474

RESULT 6
AAG06248
ID AAG06248 standard; Protein; 377 AA.
XX AC AAG06248;
XX
DT 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 2952.
XX
XX Arabidopsis thaliana protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.

PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 77.1%; Score 1942; DB 21; Length 377;
Best Local Similarity 99.7%; Pred. No. 4.4e-203;
Matches 376; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 109 MDVRSYLGGRKVLGVAQYGNLIGVTGYYTITASISLVAVGKSNCFDGHGHTADCTIS 168
Db 1 MDVRSYLGGRKVLGVAQYGNLIGVTGYYTITASISLVAVGKSNCFDGHGHTADCTIS 60
QY 169 NYPYMAVFGIIQVILSOIPNFHKLSPSLMAAVMSFTYATIGIGLAIAIVAGGKVKTSM 228
Db 61 NYPYMAVFGIIQVILSOIPNFHKLSPSLMAAVMSFTYATIGIGLAIAIVAGGKVKTSM 120
QY 229 TGTAVGVDTAAQKIWRSPQAVGDI AFAYATVLEIQTLLRSSPAENKAMKRASLVGV 288
Db 121 TGTAVGVDTAAQKIWRSPQAVGDI AFAYATVLEIQTLLRSSPAENKAMKRASLVGV 180
QY 289 STTTPFYILCGCIGYAAFGNAPGDLTDFGPPPEPWLIDFANACIAVHLIGAYQVFAQP 348
Db 181 STTTPFYILCGCIGYAAFGNAPGDLTDFGPPPEPWLIDFANACIAVHLIGAYQVFAQP 240
QY 349 IFQFVEKKNRNPDKNFITSEYVNVVPLGKFNISLFLVWRTAVVLTIVVAMIFPFF 408
Db 241 IFQFVEKKNRNPDKNFITSEYVNVVPLGKFNISLFLVWRTAVVLTIVVAMIFPFF 300
QY 409 NAILGLIGAASFPLTVYFVPMHIAQTKIKKYSARWIAKTMVCYVCLIVSLAAGSIA 468
Db 301 NAILGLIGAASFPLTVYFVPMHIAQTKIKKYSARWIAKTMVCYVCLIVSLAAGSIA 360
QY 469 GLISSVKTYKPPRTMHE 485
Db 361 GLISSVKTYKPPRTMHE 377

RESULT 7
ABB93755
ID ABB93755 standard; Protein; 481 AA.
XX
AC ABB93755;
XX

DT 31-MAY-2002 (first entry)
XX Herbicidally active polypeptide SEQ ID NO 2966.
DE Herbicidally active polypeptide SEQ ID NO 2966.
XX Herbicidally active polypeptide SEQ ID NO 2966.
XX Herbicidally active polypeptide SEQ ID NO 2966.
XX Arabidopsis thaliana.
OS Arabidopsis thaliana.
PN WO200210210-A2.
XX 07-FEB-2002.
XX 28-AUG-2001; 2001WO-EP09892.
XX 28-AUG-2001; 2001WO-EP09892.
XX (FARB) BAYER AG.
XX Tietjen K, Weidler M;
XX WPI; 2002-269010/31.
XX Identifying plant target proteins for herbicidally active compounds,
comprising aligning and comparing nucleic acid or amino acid sequences
from plant with nucleic acid or amino acid sequences from non-plant
organisms -
XX Claim 5; SEQ ID NO 2966; 26lpp + Sequence Listing; English.
XX The invention relates to identifying target proteins
(ABB90790-ABB94016) for herbicidally active compounds, comprising
aligning and comparing nucleic acid or amino acid sequences from plant
with nucleic acid or amino acid sequences from non-plant organisms using
a suitable search parameters, where plant sequences having an E-value
greater by a factor of 3 than the E-value of most similar non-plant
sequences are selected. The polypeptides or nucleic acids encoding them
are useful for identifying modulators. The identified modulators are
useful as herbicides.
XX SQ Sequence 481 AA;
Query Match 74.3%; Score 1871; DB 23; Length 481;
Best Local Similarity 73.6%; Pred. No. 3.6e-195;
Matches 345; Conservative 60; Mismatches 62; Indels 2; Gaps 2;
QY 15 ESGDAYTVSDPTKNVDEGREKRTGTWLTASAHITAVIGSVLSLAWAIAQLGWIAGTS 74
Db 11 QSFPEHEIGDTNKNFDEGRDKRTGTWMTGSAHITAVIGSVLSLAWAIAQLGWIAGTS 70
QY 75 ILLIESFTYFTSTMLADCYRAPDPVTGKRNYYTMDVRSYLGGRKVLGVAQYGNLIG 134
Db 71 VLMAFSFTYFTSTMLADCYRSPDVTGKRNYYTMDVRSYLGGRKVLGVAQYGNLIG 130
QY 135 VTVGVTITASISLVAVGKSNCFDGHGHTADCTISNYPYMAVFGIIQVILSOIPNFHKLSP 194
Db 131 ITIGVTITASISLVAVGKSNCFDGHGHTADCTISNYPYMAVFGIIQVILSOIPNFHKLSP 190
QY 195 LSIMAAVMSFTYATIGIGLAIAIVAGGKVKTSMTAGVVDVTAQKIWRSPQAVGDI 253
Db 191 LSILAAVMSFTYATIGIGLAIAIVAGGKVKTSMTAGVVDVTAQKIWRSPQAVGDI 250
QY 254 AFAYATVLEIQTLLRSS-PAENKAMKRASLVGVSTTTFFIYLCGCGIYAAFGNAPG 312
Db 251 AFAYATVLEIQTLLRSS-PAENKAMKRASLVGVSTTTFFIYLCGCGIYAAFGNAPG 310
QY 313 DFLTDGFFPEPWLIDFANACIAVHLIGAYQVFAQPIQFVEKKNRNPDKNFITSEYS 372
Db 311 NFLTGFGEYEPWLIDFANACIAVHLIGAYQVFAQPIQFVEKKNRNPDKNFITSEYS 370
QY 373 VNVPLGKFNISLFLVWRTAVVLTIVVAMIFPFFNAILGLIGAASFPLTVYFVPMH 432
Db 371 IHVPCCGDIFNLSLWRTSYVVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVVAVV 430

Query Match 59.7%; Score 1504; DB 23; Length 476;
Best Local Similarity 57.6%; Pred. No. 4,4e-155;
Matches 273; Conservative 97; Mismatches 100; Indels 4; Gaps 4;

QY 10 NSTAESGAYVSDPTKRVDEGREGKRTGMLTASAHITTVIGSGVSLIMATIQLOM 69
DB 4 NMQTVLAVD-MQOTGSKKLDDGKNNKRTGVTWTAHAHITTAIGSGVSLMATIQLOM 62
QY 70 IAGTSLILFEFTYFTSTMLADCYRAPDPVIGKRNVTMDVRSYLGKRYQLCGVAYQ 129
DB 63 LAGPVMMLFSAVTYTSTSLAACYSGDPISGKRNVTMDAARSLNGVKTLGCIYQY 122
QY 130 GNLIGVTVYTTTASISLVAAGKSNCFHDKGHTADCTISNRYMAVFGIITQVILSGIENF 189
DB 123 LNIFGVAIGYTTASISMAIKRSNCFHKSCKGKDPCHMNSNPFYMAFGIYQVILFSQIDPF 182
QY 190 HTLSFSLIMAAVSTFYATITGIGLAI-TVAGKVKGTSMGTAVGVDTAAQKTRSFQ 248
DB 183 DQLMWLISILAAVMSFTYSAGLALGIAQVYVNGKV-KGSLTGISIGA-VTEYQKIWRTRQ 240
QY 249 AVGDIAFAYAVATVLEIDTLSSPAENKAMRASLVGSTTTPFYILGCGICGIAAFGN 308
DB 241 ALGDIAPATSYSLILEIDTVASPSSEKTKATLVSVTMTMYMLCGCMGTAAFGD 300
QY 309 NMPGDELTPGFEFPEFMLIDFANACIAVHLIGAYQVPAQPIFQVFEKCCNRNYPDKKFTT 368
DB 301 LSPGNLLTGFGFYNPYMLDIAANAIVHILGAYQVYCGPLFAFIEKQASIQPPDSEFLA 360
QY 369 SEYSVNVPELGKNIISLFLVMTAVVITTVAMIFPPFNALGLIGAASFPLVYFP 428
DB 361 KQIKPIPGPKPLRLNVFRLIMRTVFVITTVISMILPFENDVGLGALGFMLTVYFP 420
QY 429 VEHMIAQTKIKKYSARMIALKTMCVCLIVSLAAGSISGLISSVYKPEPT 482
DB 421 VEMYIAQKTIKPMSTRWCLQVFSIGCLVSIAMAGSISGLVLLDKSYKPEFS 474

RESULT 10
AAG28805
ID AAG28805 standard; Protein; 464 AA.
XX
AC AAG28805;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 34161.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridization assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132486.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135153.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0138453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142820.
PR 12-JUL-1999; 99US-0142877.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.

PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	28-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match 59.6%; Score 1501.5; DB 21; Length 464;
Best Local Similarity 58.7%; Pred. No. 7.9e-155;
Matches 269; Conservative 95; Mismatches 91; Indels 3; Gaps

Qy	26	TKNVDEGREKRTGWLITASAHIIITAVIGSGVLSSLAWAIAQLGWIAAGTSLILIFSPITYP	85
	:	: : : :	:
	:	: : : :	:
Dd	7	SKYLDDDGKKRGTGSVMYTSASAHIIITAVIGSGVLSSAWATAQLGWLAPVVMLFSVVTF	66
	:	: : : :	:
	:	: : : :	:
Qy	86	TSTWLAICYRAPDPVTGKRYNTYMDVVRYSVLGRKVKOLCGVAQYNLIGTVGYTTASI	145
	:	: : : :	:
Dd	67	TSSLLAACYRSGBPISGKRNYTYMDAVRSNLGGKVTLGCIQVLYNIFGVAIGYTIASAI	126
	:	: : : :	:
	:	: : : :	:
Qy	146	SLVAVGKSNCFBPKDGHADCTISNPYMAVFGLIQVILSQIPNFHKLSPLSIMAAVMSPT	205
	:	: : : :	:
Dd	127	SMPAIRKSNCFHKSCKDPCHMNSNPMYAFGLVQLFQIOPDFDQMLSLAAVMSFT	186
	:	: : : :	:
	:	: : : :	:
Qy	206	YATIGIGLAIA-TVAGKGVKTSMTGTAVGDVDVTAQAOKIWRSPAQVGDIAPAYATVLI	264
	:	: : : :	:
Dd	187	YSSNAGLAGIAQVNVNGKV-KGSLTGISIGA-VTEQKIWRTEQALGDIAFAVSILI	244
	:	: : : :	:
	:	: : : :	:
Qy	265	EIQDTLRSSPAENKAKRPASLVGVSTTTFFYLCCGICGAAGFNNAPOGDLTDGFPEFP	324
	:	: : : :	:
Dd	245	EIQDTVKSPSEKTKWKATLVSVSVTTWPFMYLGCWGMYAAFGDLSPGNLLTGFGYPNP	304
	:	: : : :	:
	:	: : : :	:
Qy	325	WLIDPANACIAVHLIGAYOVAQPIQFQVEKKCNRYPNDKRITTSEYSVNVPLGKFNIS	384
	:	: : : :	:
Dd	305	WLIDIANAAVIHLIGAYOVYCPLFAPIEKOASIOFPDSBFIAKDIKIPIPGFKPLRN	364
	:	: : : :	:
	:	: : : :	:
Qy	385	LFLRWRTAYVWLTIVVAMIFFPNNAILGLIGAASFWPITVVPVEMHIAQTIKKYRAR	444
	:	: : : :	:
Dd	365	VFLRIWRTVEVIITVISMLLPFDNDVGLLGALGFWPLTVTFPVEMYAQKKIPWSIR	424
	:	: : : :	:
	:	: : : :	:
Qy	445	WIALKTCVYCLIVSLAAAGSIAGLISSVKTYPKPT	482
	:	: : : :	:
Dd	425	WVCILQVPSLGLVSVSAAAAAGSIAGVLLDLKSYKPFRS	462
	:	: : : :	:
	:	: : : :	:

RESULT 11
 ABB93404
 ID ABB93404 standard; Protein; 493 AA.
 AC ABB93404;
 XX ABB93404;
 XX ABB93404;
 DT 31-MAY-2002 (first entry)
 DE Herbicidally active polypeptide SEQ ID NO 2615.
 KW Herbicidal; plant; agriculture; herbicide.
 OS Arabidopsis thaliana.
 XX WO200210210-A2.
 PN

XX 07-FEB-2002.
PD
XX
XX 28-AUG-2001; 2001WO-EP09892.
PF
XX 28-AUG-2001; 2001WO-EP09892.
PR
XX (FARB) BAYER AG.
PA
XX Tietjen K, Weidler M;
PI
XX WPI; 2002-269010/31.
DR
XX Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms -
PS Claim 5; SEQ ID NO 2615; 261pp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
CC (AB90790-AB994016) for herbicidally active compounds, comprising
CC aligning and comparing nucleic acid or amino acid sequences from plant
CC with nucleic acid or amino acid sequences from non-plant organisms using
CC suitable search parameters, where plant sequences having an E-value
CC greater by a factor of 3 than the E-value of most similar non-plant
CC sequences are selected. The polypeptides or nucleic acids encoding them
CC are useful for identifying modulators. The identified modulators are
CC useful as herbicides.
XX
XX
SQ Sequence 493 AA;

Query Match 59.6%; Score 1501; DB 23; Length 493;
Best local Similarity 56.7%; Pred No. 9.8e-155;
Matches 271; Conservative 94; Mismatches 111; Indels 2; Gaps 2;

QY 8 GHNHSTAESGDAYVSDPTKNVDEDEGRKRTGTLTASAHIIITAVIGSVLSAWAIAQL 67
DB 17 GHQVFDVASHDFPPQPAKCFDDGRRLKRTGVMTASAHIIITAVIGSVLSAWAIAQL 76
QY 68 GWIAGTSLILISFTYFTFSTMLADCYRAPDVTGKRNVTYMDVRSYLGGRVQLGVA 127
DB 77 GWIAGPAVMLLSLVLTYSSTLSDCYRTGDVAVSGKRNVTYMDVAVRSILGFFKXICGLI 136
QY 128 QYGNLIGVAVGTITATISLVANGKSCNCPDKHTADCTISNPMVAAGIIIVILISQIP 187
DB 137 QYINLFGAIGTITASISMMALKRSCNCPDKSGKDPCHMSSNPMIVFGVAIILISQVP 196
QY 188 NPHKLSFLSMAVMSFTATIGIATVAGGKYGKTSMTGTAVGVDTAAQKIMRSF 247
DB 197 DDDQIMWISIVAAVMSFTYSAIGLAGIYQVANGVFKGSLTGISIGT-VTQOKIMRIF 255
QY 248 QAVGDIAPYAVATVLIETIOTLRSSPAENKAMKASLVGVSTTFEPIYILGCGIYAAFG 307
DB 256 QALGDIAPYAVSVVLIETIOTLRSSPAESKTKKATKISIAVTTIFMYLCSMGYAAFG 315
QY 308 NNAPGDFLDFGFEFBEFMLIDFANACIAVHLIGAXOVFAOPRFOVEKKCNRYPNKFI 367
DB 316 DDAPEGLLTGFGYNPFLDIDANAIVLHVGAYOVFAPIFAFLKSYAERYPNDFL 375
QY 368 TSEYVAVP-FLGKFNISLFLVMTAYVYITVVMIFPFENAIIGLIGASFWPLTVY 426
DB 376 SKFEFLRIGFSPYKVNFRMYRSGFVYTTIVISLMMPFNDVCGIIGALGFWPLTVY 435
QY 427 FPEVMEIAOTKKIKKYSARWIALKTCVCLIVSLAAAGSIAGLISSVKTYKPFRTMH 484
DB 436 FPEVEMIKORKEKSTRWVLCMLSVACIIVSVAGVGSIGAVMDLKVYKEFKSTY 493

RESULT 12
AAG42351
ID AAG42351 standard; Protein; 466 AA.
XX

AC AAG42351;
XX
XX 18-OCT-2000 (first entry)
DT
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 52809.
DE
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132484.
XX 06-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132863.
XX 14-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134376.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 17-JUN-1999; 99US-0139453.
XX 18-JUN-1999; 99US-0139492.
XX 18-JUN-1999; 99US-0139454.
XX 18-JUN-1999; 99US-0139455.
XX 18-JUN-1999; 99US-0139456.
XX 18-JUN-1999; 99US-0139457.
XX 18-JUN-1999; 99US-0139458.
XX 18-JUN-1999; 99US-0139459.
XX 18-JUN-1999; 99US-0139460.
XX 18-JUN-1999; 99US-0139461.
XX 18-JUN-1999; 99US-0139462.

Query Match 59.3%; Score 1495; DB 21; Length 466


```

Qy 267 QDTRSSPAENKAMKASLVGVSTTFFPYLLGCGIGYAAFGNNAPEDFLTDFGFEPEPM 326
Db 248 QDTRSSPAENKAMKASLVGVSTTFFPYLLGCGIGYAAFGNNAPEDFLTDFGFEPEPM 307
Qy 327 IDPANACIAVHLIGAAYOAPRPFQFEKKCNRPNDKFTSEYGVNP-FLGKNISL 385
Db 308 LDVANAIAIYHLVGAAYOAPRPFQFEKKCNRPNDKFTSEYGVNP-FLGKNISL 367
Qy 386 FLVWRTAVVITTVVAMTFPFENALIGLGAASFWPLTVYFPEVNHIAQTKIKKYSARW 445
Db 368 FRAVYRSGFVVLTVYSMLMPFENDVVGILGALGFPLTVYFPEVNHIAQTKIKKYSARW 427
Qy 446 IALKTMCVCLIVSLAAGSINAGLISVYKYPFRTH 484
Db 428 VCLQMLSCGLMTITVAVGVSIGVMDLVKVKPFKTTY 466

RESULT 13
ID ABB93937 standard; Protein; 466 AA.
AC ABB93937;
XX
XX 31-MAY-2002 (first entry)
XX
XX Herbicidally active polypeptide SEQ ID NO 3148.
XX
XX Herbicidal; plant; agriculture; herbicide.
XX
XX Arabidopsis thaliana.
XX
XX WO200210210-A2.
XX
XX 07-FEB-2002.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX 28-AUG-2001; 2001WO-EP09892.
XX
XX (FARB ) BAYER AG.
XX
XX Tietjen K, Weidler M;
XX
XX WPI; 2002-269010/31.
XX
XX Identifying plant target proteins for herbicidally active compounds,
XX comprising aligning and comparing nucleic acid or amino acid sequences
XX from plant with nucleic acid or amino acid sequences from non-plant
XX organisms -
XX
XX Claim 5; SEQ ID NO 3148; 261dp + Sequence Listing; English.
XX
XX The invention relates to identifying target proteins
XX (ABB90790-ABB94016) for herbicidally active compounds, comprising
XX aligning and comparing nucleic acid or amino acid sequences from plant
XX with nucleic acid or amino acid sequences from non-plant organisms using
XX suitable search parameters, where plant sequences having an E-value
XX greater by a factor of 3 than the E-value of most similar non-plant
XX sequences are selected. The polypeptides or nucleic acids encoding them
XX are useful for identifying modulators. The identified modulators are
XX useful as herbicides.
XX
XX Sequence 466 AA;
XX
XX Query Match 59.3%; Score 1495; DB 23; Length 466;
XX Best Local Similarity 58.0%; Pred. No. 4.1e-154;
XX Matches 266; Conservative 98; Mismatches 93; Indels 2; Gaps 2;
Qy 27 KAVDEGKREKRTGTWLTASHIITAVVGSVLSAAWIAQLGWTAGSIIILISFTTYPT 86
Db 9 KCFDDGRLKRSQVWTAASHIITAVVGSVLSAAWIAQLGWTAGSIIILISFTTYPT 68
Qy 87 STMADCYRAPDPVTGKNTYMDVVRSYLGRKVLGCGVAGQGNLIGTVGTITASIS 146

```

```

Db 69 STLLSDCYRTGPFVSGKRYTYMDAVRSILGGRFKICGLIYLNFGITVGTITASIS 128
Qy 147 LVAAGSNCFHDKGHTADCTISNYPMAVAGLIIQVLIISQIPNPKLSFISIMAAVSEFY 206
Db 129 KMAIKRSCNFHESGCKNCPCHMSNPYIMFVTEILISQIKDPDQIWMLSIVAAISFTY 188
Qy 207 ATIGLAIATVAGGKVGKTSMTGTAAGVDTAAQKIRWSFOAVGDIAPAYATVLIET 266
Db 189 SAIGLALGIQVAAANGVAVGSLTGISIGA-VTQTKIWTFTFQLDGIAPAYSVVLIEI 247
Qy 267 QDTRSSPAENKAMKASLVGVSTTFFPYLLGCGIGYAAFGNNAPEDFLTDFGFEPEPM 326
Db 248 QDTRSSPAENKAMKASLVGVSTTFFPYLLGCGIGYAAFGNNAPEDFLTDFGFEPEPM 307
Qy 327 IDPANACIAVHLIGAAYOAPRPFQFEKKCNRPNDKFTSEYGVNP-FLGKNISL 385
Db 308 LDVANAIAIYHLVGAAYOAPRPFQFEKKCNRPNDKFTSEYGVNP-FLGKNISL 367
Qy 386 FLVWRTAVVITTVVAMTFPFENALIGLGAASFWPLTVYFPEVNHIAQTKIKKYSARW 445
Db 368 FRAVYRSGFVVLTVYSMLMPFENDVVGILGALGFPLTVYFPEVNHIAQTKIKKYSARW 427
Qy 446 IALKTMCVCLIVSLAAGSINAGLISVYKYPFRTH 484
Db 428 VCLQMLSCGLMTITVAVGVSIGVMDLVKVKPFKTTY 466

RESULT 14
ID AAG13704 standard; Protein; 493 AA.
AC AAG13704;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO. 13297.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0123825.
XX
XX 05-MAR-1999; 99US-0123180.
XX
XX 09-MAR-1999; 99US-0123548.
XX
XX 23-MAR-1999; 99US-0125788.
XX
XX 25-MAR-1999; 99US-0126264.
XX
XX 29-MAR-1999; 99US-0126785.
XX
XX 01-APR-1999; 99US-0127462.
XX
XX 06-APR-1999; 99US-0128234.
XX
XX 08-APR-1999; 99US-0128714.
XX
XX 16-APR-1999; 99US-0129845.
XX
XX 19-APR-1999; 99US-0130077.
XX
XX 21-APR-1999; 99US-0130449.
XX
XX 23-APR-1999; 99US-0130510.
XX
XX 23-APR-1999; 99US-0130891.
XX
XX 28-APR-1999; 99US-0131449.
XX
XX 30-APR-1999; 99US-0132048.
XX
XX 30-APR-1999; 99US-0132407.
XX
XX 04-MAY-1999; 99US-0132484.
XX
XX 05-MAY-1999; 99US-0132485.
XX
XX 06-MAY-1999; 99US-0132486.
XX
XX 06-MAY-1999; 99US-0132487.
XX
XX 07-MAY-1999; 99US-0132863.
XX
XX 11-MAY-1999; 99US-0134256.

```

PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159285.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.

PR 22-OCT-1999; 99US-0160989.
 PR 25-OCT-1999; 99US-0161404.
 PR 25-OCT-1999; 99US-0161405.
 PR 25-OCT-1999; 99US-0161406.
 PR 26-OCT-1999; 99US-0161359.
 PR 26-OCT-1999; 99US-0161360.
 PR 26-OCT-1999; 99US-0161361.
 PR 28-OCT-1999; 99US-0161920.
 PR 28-OCT-1999; 99US-0161921.
 PR 28-OCT-1999; 99US-0161922.
 PR 28-OCT-1999; 99US-0161923.
 PR 29-OCT-1999; 99US-0162142.

Query Match 59.2%; Score 1492; DB 21; Length 493;
 Best Local Similarity 56.5%; Pred. No. 9, 5e-154;
 Matches 270; Conservative 94; Mismatches 112; Indels 2; Gaps 2;

QY 8 GHNHTASGDVYTSDFPKNVDEGRKRTGTWLTASNHITTAIVIGSVLSLAMAIALD 67
 DB 17 GHQVDVASHDVPPOPAKPCPDGGRKRTGTWLTASNHITTAIVIGSVLSLAMAIALD 76
 QY 68 GMIAGTSILIFSTFTYFTSTMLADCYRAPDPVTKRNTYMDVRSYLGKRVOLCGVA 127
 DB 77 GMIAPAVMLFSLVTLVSTLSDCYRTGDVSGKNTYMDVARSILGFKFKICGLI 136
 QY 128 QYGNLIGVGYTTTASISLVAVSKNCFHDKHTADCTISNYPYMAVFGIIOVILSQIP 187
 DB 137 QYLNLFVIAIGYTTAASISMAIKRSCNCFHSGKOPCHMSNPYMAVFGVAILLSQVP 196
 QY 168 NFHKLSFISMAAVNSFTYATIGLAIATTVAGKVKTSMTGTAVGVDTAAOKTMSF 247
 DB 197 DFDQIMWISIVAAVNSFTYATIGLAIATTVAGKVKTSMTGTAVGVDTAAOKTMSF 255
 QY 248 QAVGDIAPAVATVLEIOTDTRSSPAENKAMKRASLVGSTTTFYILCGCIGYAAG 307
 DB 256 QALGDIAPAVATVLEIOTDTRSSPAENKAMKRASLVGSTTTFYILCGCIGYAAG 315
 QY 308 NNAPGDLTDFGFPFWLIDPANACIAVHLIGAYQVFAQPIFOVEKKCNRYPDNKEI 367
 DB 316 DAAPNLLTDFGFPFWLIDPANACIAVHLIGAYQVFAQPIFOVEKKCNRYPDNKEI 375
 QY 368 TSEYVNVNPFEGKRNISLFLVMTAVYVITTVAMTFPFNNAIGLIGASFWPLTY 426
 DB 376 SKEFIRIPGFSKYVNVFMYVRSFVVTTVISMMPFNDVVGILGALGFMDLTY 435
 QY 427 PVEVNHIOTRIKKYSARIALKTMQVCLIVSLAAGSINGLISSVKTYKPERFMH 484
 DB 436 PVEVNHIOTRIKKYSARIALKTMQVCLIVSLAAGSINGLISSVKTYKPERFMH 493

RESULT 15
 ABB91267
 ID ABB91267 standard; Protein; 476 AA.
 AC ABB91267;
 XX
 DT 31-MAY-2002 (first entry)
 XX
 DE Herbicidally active polypeptide SEQ ID NO 478.
 XX
 KW Herbicidal; plant; agriculture; herbicide.
 XX
 OS Arabidopsis thaliana.
 XX
 FN WO200210210-A2.
 XX
 PD 07-FEB-2002.
 XX
 PF 28-AUG-2001; 2001WO-EP09892.
 XX
 PR 28-AUG-2001; 2001WO-EP09892.
 XX
 PA (FARB) BAYER AG.
 XX

PI Tietjen K, Weidler M;
 XX WPI; 2002-269010/31.
 DR
 XX
 PT Identifying plant target proteins for herbicidally active compounds,
 PT comprising aligning and comparing nucleic acid or amino acid sequences
 PT from plant with nucleic acid or amino acid sequences from non-plant
 PT organisms -
 XX
 PS Claim 5; SEQ ID NO 478; 261bp + Sequence Listing; English.
 XX
 CC The invention relates to identifying target proteins
 CC (ABB90790-ABB94016) for herbicidally active compounds, comprising
 CC aligning and comparing nucleic acid or amino acid sequences from plant
 CC with nucleic acid or amino acid sequences from non-plant organisms using
 CC suitable search parameters, where plant sequences having an E-value
 CC greater by a factor of 3 than the E-value of most similar non-plant
 CC sequences are selected. The polypeptides or nucleic acids encoding them
 CC are useful for identifying modulators. The identified modulators are
 CC useful as herbicides.

Query Match 55.8%; Score 1406.5; DB 23; Length 476;
 Best Local Similarity 55.8%; Pred. No. 2e-144;
 Matches 256; Conservative 91; Mismatches 101; Indels 11; Gaps 4;

QY 30 DEDGREKRTGTWLTASNHITTAIVIGSVLSLAMAIALDQMIAGTSILIFSTFTYFTSM 89
 DB 21 DDDGRKRTGTWLTASNHITTAIVIGSVLSLAMAIALDQMIAGTSILIFSTFTYFTSM 80
 QY 90 LADCYRAPDPVTKRNTYMDVRSYLGKRVOLCGVAGNLIIGVTVTTTASISLVA 149
 DB 81 LSCYRSGSVTKRNTYMDVRSYLGKRVOLCGVAGNLIIGVTVTTTASISLVA 139
 QY 150 VKSNCFHDKHTADCTISNYPYMAVFGIIOVILSQIPNFHKLSFISMAAVNSFTYAT 209
 DB 140 ---TSCQKMGNDPCHVNGVNYMIAFGIVQIIFSQIPDQMLMISIVAAVNSFAYSAT 196
 QY 210 GIGLAIATVAGKVKTSMTGTAVG-----DYTAOKTMSFQAVGDIAPAVATVLI 264
 DB 197 GIGLGVSKVENEKIKGSLGTGTVTLGVTSSQKIMRTFOSIGNIAFAVYSMILI 256
 QY 265 EIOTDTRSSPAENKAMKRASLVGSTTTFYILCGCIGYAAGNNAPGDLDFGFPF 324
 DB 257 EIOTDTRSSPAENKAMKRATFVSAVTTVFYMLCGCVGIAAGDNNAPGDLAHGFRNRY 316
 QY 325 WLIDPANACIAVHLIGAYQVFAQPIFOVEKKCNRYPDNKEITSEYVNVFELGK-FNI 383
 DB 317 WLIDPANALAIIVHIGAYVYQPLFAFVEKASRRFPSESEVTKIKIOL-FPGKPFNL 375
 QY 384 SLFLVMTAVYVITTVAMTFPFNNAIGLIGASFWPLTYFPVEVNHIOTRIKKYS 443
 DB 376 NLFRLVMTAVYVITTVAMTFPFNNAIGLIGASFWPLTYFPVEVNHIOTRIKKYS 435
 QY 444 RWIALKTMQVCLIVSLAAGSINGLISSVKTYKPERFMH 482
 DB 436 KWCLOVLSVTLFVSAAGSIVGIVSDLVKTYKPERFQS 474

Search completed: November 22, 2002, 13:29:57
 Job time : 41 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using SW model

Run on: November 22, 2002, 13:30:02 ; Search time 12 Seconds

(without alignments)
632.985 Million cell updates/sec

Title: US-09-854-562-2

Perfect score: 2519
Sequence: 1 MKSFTEGHMSTAESGDY.....SIAGLISSEVTKYKPFRTWHE 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100480 seqs, 15661496 residues

Total number of hits satisfying chosen parameters: 100480

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : Published Applications_AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/BCT_NEW_PUB pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	388	15.4	468	10	US-09-860-232A-8
2	212.5	8.4	88	10	US-09-860-232A-18
3	150.5	6.0	504	10	US-09-918-497-67
4	143.5	5.7	547	10	US-09-789-919-66
5	130.5	5.2	547	10	US-09-776-705-4
6	128	5.1	525	10	US-09-871-503-2
7	128	5.1	525	10	US-09-940-919-2
8	124.5	4.9	547	10	US-09-776-705-2
9	122.5	4.9	547	10	US-09-860-232A-5
10	112.5	4.5	506	10	US-09-776-705-5
11	111.5	4.4	619	10	US-09-741-153-2
12	107	4.2	440	10	US-09-815-242-13429
13	104	4.1	1212	9	US-09-981-353-16
14	101.5	4.0	551	9	US-09-895-913A-268
15	96	3.8	2516	10	US-09-817-514A-2
16	95.5	3.8	326	10	US-09-815-242-13409
17	95	3.8	411	10	US-09-815-242-5128
18	95	3.8	462	10	US-09-815-242-10161
19	94.5	3.8	284	10	US-09-800-103-8

20	94.5	3.8	284	10	US-09-800-103-20	Sequence 20, Appl
21	94.5	3.8	310	10	US-09-800-103-30	Sequence 30, Appl
22	94.5	3.8	481	10	US-09-815-242-10979	Sequence 10979, A
23	93	3.7	542	10	US-09-801-368-48	Sequence 48, Appl
24	92.5	3.7	680	10	US-09-996-194-10	Sequence 10, Appl
25	91	3.6	441	10	US-09-950-071-2	Sequence 2, Appl
26	91	3.6	482	10	US-09-815-242-12941	Sequence 12941, A
27	91	3.6	619	10	US-09-741-153-4	Sequence 4, Appl
28	89.5	3.6	290	10	US-09-800-103-10	Sequence 10, Appl
29	89.5	3.6	290	10	US-09-800-103-22	Sequence 22, Appl
30	89.5	3.6	463	10	US-09-815-242-10144	Sequence 10144, A
31	89.5	3.6	619	10	US-09-800-065-2	Sequence 2, Appl
32	89	3.5	632	10	US-09-800-103-4	Sequence 4, Appl
33	89	3.5	392	10	US-09-800-103-16	Sequence 16, Appl
34	89	3.5	418	10	US-09-800-103-26	Sequence 26, Appl
35	89	3.5	430	10	US-09-800-103-12	Sequence 12, Appl
36	89	3.5	430	10	US-09-800-103-24	Sequence 24, Appl
37	89	3.5	456	10	US-09-800-103-34	Sequence 34, Appl
38	88.5	3.5	477	10	US-09-815-242-13840	Sequence 13840, A
39	88	3.5	466	10	US-09-815-242-13831	Sequence 13831, A
40	87.5	3.5	247	10	US-09-886-468-24	Sequence 24, Appl
41	87.5	3.5	289	10	US-09-815-242-5129	Sequence 5129, Ap
42	87.5	3.5	438	10	US-09-815-242-11824	Sequence 11824, A
43	87.5	3.5	658	10	US-09-815-242-12272	Sequence 12272, A
44	87	3.5	484	10	US-09-815-242-5660	Sequence 5660, Ap
45	86	3.4	439	10	US-09-815-242-5660	Sequence 5660, Ap

ALIGNMENTS

```
RESULT 1
US-09-860-232A-8
; Sequence 8, Application US/09860232A
; Patent No. US2002028494A1
; GENERAL INFORMATION:
; APPLICANT: Curtiss, Roy A.J.
; TITLE OF INVENTION: 57256 AND 58289, NOVEL HUMAN
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; FILE REFERENCE: 381552001500
; CURRENT APPLICATION NUMBER: US/09/860.232A
; PRIOR FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/205,288
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; TYPE: PRT
; LENGTH: 468
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Consensus amino acid sequence
US-09-860-232A-8

Query Match      15.4% Score 388; DB 10; Length 468;
Best Local Similarity 30.4% Pred. No. 2.5e-28;
Matches 139; Conservative 67; Mismatches 183; Indels 68; Gaps 20;

QY 66 QLGWLAGTSLILIFSPIT--YFTSTMLADCYRAP--DPTGKRN--YTYMDVVR--SYLGG 118
   |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 17 QLGWGLVLLVLLVGRITLYGHFTGLLVKCYEEBGEYVPGKREKGSYLDLGRQAYGG 76

QY 119 RKYQLCGV---AQGNLIGTVGVGTTTASISLVAVGK--NCF-----HDKGHTAD--C 165
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 77 KGLLTSFVGTGVQVYVNLPGVNIYGILADGLLPKLIISFLLIQVDNNGVSDIGC 136

QY 166 -TISNVPYAVF--GIQVLTISOIPNFKHS-----PLSIMAAMSFTY-ATIGIG 212
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 137 DLISGSMWIIITPAIITPLSTIPAFNLISAGNLISLVIISSISAFSSLAYISITSEI 196

QY 213 LAIATVAG-----GKVGKTSMTGTAAGVDVTAOKIWRSFQAVGDIAFAVAATVLIIEIQ 267
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 197 IYVAIVLAGIFVLLGAVYKILMSVETLAVVPSVTKUTGFLAIGIIVFAFEGHVALVLPQ 256
```



```

QY 105 -----NYTMDVRSYLG----- 117
DB 146 FQMPGKIGAFISITMONIGAMSSYLFIIKXELPEVIRAFMGJEENTGEWYINGNYLVLFV 205
QY 118 --GRVQVLCGVAQGNLGVTVGYTTTASISLVAVGSKNCF-----HDKHTIAC 165
DB 206 SVGIILPLSLKNLGYL-GYTSGPSLSQVFPVSVIYKKFIPCPPLADHNNG---NL 261
QY 166 TISNRYMAVFGIIOVILISQIPNFHLSISMAAVSFTYATIGLAIATVAGKVC 225
DB 262 TFNN-----TLPIHMSLPNDSSSGVNF---MDVAHNN-PAGDEKQVAG----- 304
QY 226 TSMGTAVGVDTAAQKI-----RSFOVAGDIAFAVAVATVLEIQTDLRSSPAEN 277
DB 305 -PLHNGVEYEQAQAKCQPKYFVFNSTRAYAIPIIAFAVCGPEVLPIYSEKD--RSR 361
QY 278 KMKRSASLVGVSTTFEYLLCGCIGYAAFGNNAPOGFLDFFGFEPFMLIDPANACIAVH 337
DB 362 RKMQTVSNISISGMLVMYLLALFGYLSFYGDVEDELLHAYS---KYTFDTALLMVR 416
QY 338 LIGAYOV-FAQIPQFVEKKCNRYPDNKFITSEYSVNVF--LGKFNISLFPILVWRTAY 394
DB 417 LAVVAVTLTVIVLF-----PIRTSVITLFPKRPFSWLKAFGLA-----ATI 460
QY 395 VITTVVAMIPFPFNAILGLIGAASFMPPLTVYFPEVMHIAQTKIKKYSARWIALKTMCYV 454
DB 461 IALNNILVILVPTIKYIFGFIGASSATMILFILPAFYL--KLVKKEPLRSPOKIGALV 517
QY 455 CLIVSLAAGSIAGLI 471
DB 518 FLVTGIIFFMGSMALII 534

```

RESULT 5

```

US-09-776-705-4
; Sequence 4, Application US/09776705
; Patent No. US2002082191A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; FILE REFERENCE: CLO01010
; CURRENT APPLICATION NUMBER: US/09/776,705
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Rattus norvegicus
US-09-776-705-4

```

```

Query Match 5.2%; Score 130.5; DB 10; Length 547;
Best Local Similarity 17.9%; Pred. No. 0.00024;
Matches 92; Conservative 87; Mismatches 215; Indels 119; Gaps 16;

```

```

QY 32 DGRBKRTGWLTAHAHTTAIVIGSVLSLMAIOLGMIAGTISILIFSTYFTSTMLA 91
DB 68 DEHHQGTISFGMSFPLSNAINMSGILGLSYAMANTGIVLFVIMLTVAIILSYVHLL 127
QY 92 DCYRAPDPVTGKR-----NYTMDVRSYLG- 117
DB 128 KTAAGGGLIYERLGEKAFGMPGKIGAFISITMONIGAMSSYFIITYELPEVIRAFMG 187
QY 118 -----GRVQVLCGVAQGNLGVTVGYTTTASISLVAVGSKNCF- 156
DB 188 EENTGEWYINGNYLVLFVSVGIILPLSLKNLGYL-GYTSGPSLSQVFPVSVIYKKRQ 246
QY 157 -----HDKHTIADCTISNRYMAVFGIIOVILISQIPNFHLSISMAAVSFTYA 207
DB 247 ICPPLPVLADHNNG---NLTFNNTLPMHVI-----MLPNNSSESTGMNFM--VDYTHR 292

```

```

QY 208 TIGIGLAIATVAGKVCSTMTGTAAGVDTAAQKI-----RSFOVAGDIAFAVAY 259
DB 293 D-PEGDEKPAAG-----PLHSGVEYEAHSGDKCQPKYFVFNSTRAYAIPIIAFAV 345
QY 260 ATVLEIQTDLSSPAENKAMKRSASLVGVSTTFEYLLCGCIGYAAFGNNAPOGFLDFFG 319
DB 346 HPEVLPYSELDD--RSRKMQTVSNISITGMLVMYLLALFGYLSFYGEVEDELLHAYS 403
QY 320 FPEFPLIDFANACIAVHLIGAYOV-FAQIPQFVEKKCNRYPDNKFITSEYSVNVFPL 378
DB 404 ----KYTFDTALLMVRILAVVAVTLTVIVLF-----PIRTSVITLFPKRPFS 449
QY 379 GKFNISLFPILVWRTAYVITTVVAMIPFPFNAILGLIGAASFMPPLTVYFPEVMHIAQTKI 438
DB 450 WKHFGIAIIL-----IALNNVILVILVPTIKYIFGFIGASSATMILFILPAFYL--XL 501
QY 439 KYTSARWIALKTMCYVCLIVSLAAGSIAGLI 471
DB 502 VKKEPLRSPOKIGALVFLVTGIIFFMGSMALII 534

```

RESULT 6

```

US-09-871-503-2
; Sequence 2, Application US/09871503
; Patent No. US20020076758A1
; GENERAL INFORMATION:
; APPLICANT: George Christian Terstappen
; APPLICANT: Cinzia Felicitia Sala
; TITLE OF INVENTION: POLYPEPTIDE
; FILE REFERENCE: QG 1013
; CURRENT APPLICATION NUMBER: US/09/871,503
; CURRENT FILING DATE: 2001-05-31
; PRIOR APPLICATION NUMBER: GB 0013239.9
; PRIOR FILING DATE: 2000-06-01
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-871-503-2

```

```

Query Match 5.1%; Score 128; DB 10; Length 525;
Best Local Similarity 19.8%; Pred. No. 0.0004;
Matches 100; Conservative 79; Mismatches 228; Indels 98; Gaps 20;

```

```

QY 7 EGHNSTAESGDAYTVSDPTKNV---DEGGRKRTGWLTAHAHTTAIVIGSVLSLMA 62
DB 80 EGDHYQSGGAPLPPSGSKDQVGGGEGFGHDKPKITAMEAGMVTNMIQGMFVLGLPY 139
QY 63 ATAOLGMIAGTISILIFSTYFTSTMLADY--RAPDPVTGKRNTYTDVRS----- 114
DB 140 AILHGGYL-GLFLIIFAAVCCYTGKILILACYLENEDDEVVRVNDSYAIAANACCAPRF 198
QY 115 -YLGRKQVLCGVAQGNLGVTVGYTTTASISLVAVGSKNCFHDKHTIADCTISNRYPM 173
DB 199 PLUGGRVY--NVAQIIEIVMTCILYVV-----SGNLMNVSFPL 236
QY 174 AV-----QVILISQIPNFHLSISMAAVSFT---TYATIGLAIATVAGKVC 223
DB 237 PVSQKSMSTIAATVALL-----PCAFLEKTKAVSFFSLCTLAHFVNIIVIAVCLSLA 289
QY 224 GKTSMTGTAAGVDTAAQKIRSFQ-AVGDIAFAVAVATVLEIQTDLSSPAENKAMK 282
DB 290 RDWAMEKVFYIDV-----KKFPISIGIIVFYSYSLQIFLSLSEGNMQ-OPSEFPCMN 341
QY 283 ASLVGVSTTFEYLLCGCIGVA-----AFGNNAPOGFLDFFGFEPFMLIDFANACIAVH 337
DB 342 WTHIACVLKGFLAVVAYITMADETKEVITDNLPGSIRAVNIF---LVAQLLSYPLP 397
QY 338 LIGAYOVFAQIPQFVEKKCNRYPDNKFITSEYSVNVFPLGKFNISLFPILVWRTAYVVI 397

```

Db 398 FFAAEVLEKSLFQ-----EGSRAPPACYS-----GDGRKLSWGLTLRCALVVF 442
Qy 398 TTVVAMIFPPFNAILGLIGAASFWPLTVVFPVEMHIAQTAKIKYSARW-----IALKTM 451
Db 443 TLIMAIYVPHFALLMGLTSLTGAGLCFLPLSLFHL---RLWRKLLWHQVFPDVAIFVI 499
Qy 452 CVVCLIVSLLAAAGSTAGLISVKT 476
Db 500 GGIC---SVSGFVHSLGLEIAVRT 521

RESULT 7

US-09-940-919-2
; Sequence 2, Application US/09940919
; Patent No. US20020082390A1
; GENERAL INFORMATION:
; APPLICANT: Friddle, Carl Johan
; APPLICANT: Gerhardt, Brenda
; APPLICANT: Hu, Yi
; TITLE OF INVENTION: No. US20020082390A1el Human GABA Transporter Protein and Polynucleotide
; FILE REFERENCE: Same
; CURRENT APPLICATION NUMBER: US/09/940,919
; CURRENT FILING DATE: 2001-08-28
; PRIOR APPLICATION NUMBER: US 60/230,178
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 525
; TYPE: PRT
; ORGANISM: homo sapiens
US-09-940-919-2

Query Match 5.1%; Score 128; DB 10; Length 525;

Best Local Similarity 19.8%; Pred. No. 0.0004;

Matches 100; Conservative 79; Mismatches 228; Indels 98; Gaps 20;

Qy 7 EGNHSTASGDAYTVSDTKNV-----DEGREKRTGTLWTASAHITAVIGSGVLSLAW 62
Db 80 EGDIIHQSGAPLPSPSGSKDQVGGGEGFGHDKPITAEAGWNTNAIQGFVLGLPY 139
Qy 63 ALAQLGWIAGTSLILFFSIFITYTSLMADCY--RAPDPVTGKNTYMDVVR-----114
Db 140 ALHGGYL-GLFIIIPAAVCCVTGKILIACLYEENDEGEVVRVRSYVAIANACAPRF 198
Qy 115 -YLGRKVLQCGVAQYGNLIGTVGTITASISLAVGKSNCFPHDKGHTADCTISNYPYM 173
Db 199 PTLGGRVV--NVAQIIEIEMTCILVVV-----SGNLMYNSFPGL 236
Qy 174 AV-----FGII--QVILSQIPNFHKLSELSIMAAVMSF-----TVATIGIGLAIATVAGGV 223
Db 237 PVSQKSWSIATAVLL-----PCAFKLNKAVSKFSLLCTLAHFVINILVIAYCLSRA 289
Qy 224 GXTSMTGTAVGDVDTAAQIKRSFQ-AVGDIAPAYAYATVLIBIOTLRSSPAENKAMKR 282
Db 290 RQWAEKVFYIDV-----KKPDISIGLIVFSYTSQIFLSLEGNMQ-QPSEFHCMMN 341
Qy 283 ASLVGVSTTTFFYLCCGICGYA-----AFGNAPGDFLTDGFFFPFLLIDFANACIAVH 337
Db 342 WTHIAACVLKGLPALVAYLTWADETKEVITDNLPGSIRAVVNIF-----LVAKALLSYPLP 397
Qy 338 LIGAVQVFAQPIQFQVEKKCNRNYPDNKFTITSEYSNVFPFLGFKNLSLRLVWRAYVVI 397
Db 398 FFAAEVLEKSLFQ-----EGSRAPPACYS-----GDGRKLSWGLTLRCALVVF 442
Qy 398 TTVVAMIFPPFNAILGLIGAASFWPLTVVFPVEMHIAQTAKIKYSARW-----IALKTM 451
Db 443 TLIMAIYVPHFALLMGLTSLTGAGLCFLPLSLFHL---RLWRKLLWHQVFPDVAIFVI 499
Qy 452 CVVCLIVSLLAAAGSTAGLISVKT 476
Db 500 GGIC---SVSGFVHSLGLEIAVRT 521

RESULT 8

US-09-776-705-2
; Sequence 2, Application US/09776705
; Patent No. US20020082191A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001010
; CURRENT APPLICATION NUMBER: US/09/776,705
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Human
US-09-776-705-2

Query Match 4.9%; Score 124.5; DB 10; Length 547;

Best Local Similarity 17.7%; Pred. No. 0.00088;

Matches 98; Conservative 90; Mismatches 191; Indels 174; Gaps 20;

Qy 32 DGRKRTGTLWTASAHITAVIGSGVLSLAWIAQLGWIAGTSLILFFSIFITYTSTMLA 91
Db 68 DEHPGTTSPFGMSFNLGSLVAMAYTVGLFIIMLLAVAILSLYSVHLL 127
Qy 92 DCYRAPDPVTGKR-----NYTYMDVVRSLG- 117
Db 128 KTAKEGGLIYEKLEKAFGWPCKIGAFVITMGNIGAMSSVLPFIKPELPIRAFML 187
Qy 118 -----GRKVLQCGVAQYGNLIGTVGTITASISLAVGKSNCFH 157
Db 188 BENTGEWLNGLYLIIFVSUGIILPLSLKLNGLYL-GYTSGLSLTCWVFFSV---VIY 242
Qy 158 DKGH-----TADCTISNYPYMAVEGIIQVILSQIPNFHKLSELSIMAAVMSFTYATIGI 211
Db 243 KKFQIPCLPLVLHDHVGNLSPFNNTLPMHVVML---PNNSESDVNFV---MDYTHRN-PA 295
Qy 212 GLAIATVAGGVKGTSMGTAVGVDTAAQKI-----WRSFOAVGDIAPAYAVATVL 263
Db 296 GL-----DENQAKSLHSDSVEYEAHSDDKCEPKYFVFNSTAYAIPLVFAFVCHPEV 349
Qy 264 IBIQDILRSSPAENKAMKRASLVGVSTTTFFYLCCGICGYAAGFNNAFGDFLTDGFFEP 323
Db 350 LPIYSELKD--RSRRKQTVSNISITGMVLYLAALFGY-----LTFYGEVED 396
Qy 324 FWLIDFANACTIAVHLIGAYQVPAQPIQFQVEKKCNRNYPDNKFTITSEYSNVFPFLGKPMI 383
Db 397 -----ELHAY-----SKVYTLDPILL-MVRL 417
Qy 384 SLFRLVWRATVAVVI---TTVVAMIFP-----PFNA-----I 411
Db 418 AVLVAVTQVPIVLPIRTSVITLLFPKPPSWIRHFLIAAVLIANNVVLVLTPIKVI 477
Qy 412 LGLIGAASFPLTVVFPVEMHIAQTAKIKYSARWIALKTMCYVCLIVSLLAAAGSTAGI 471
Db 478 FGFIGASSATMLIFILPAVFYKLVKKETFRS---PQKVGALIFLVVGGIFMFGMSA-LI 533
Qy 472 SSVKTYKPPRTWH 484
Db 534 IIDWIYDPPNSKH 546

RESULT 9

US-09-860-232A-5
; Sequence 5, Application US/09860232A
; Patent No. US2002002849A1
; GENERAL INFORMATION:
; APPLICANT: Curtis, Rofy A.J.


```

; TITLE OF INVENTION: 57256 AND 58289, NOVEL HUMAN
; TITLE OF INVENTION: TRANSPORTERS AND USES THEREOF
; FILE REFERENCE: 38152001500
; CURRENT APPLICATION NUMBER: US/09/860,232A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: US 60/205,288
; PRIOR FILING DATE: 2000-05-19
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 547
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-860-232A-5

Query Match      4.9%; Score 122.5; DB 10; Length 547;
Best Local Similarity 18.0%; Pred. No. 0.0014;
Matches 94; Conservative 90; Mismatches 226; Indels 112; Gaps 18;

QY 32 DDERKRTGTLTASAHITAVIGSGVLSLAMAIAQLGWIAGTSLILFSPITFTSTMLA 91
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 68 DEHHPTTSFGMSFNLNAIMSGSILGLSYAMANTGILLFTMLLAVALLSLYSVHLL 127
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

QY 92 DQYRAPDPVTGK-----NYTMDVRSYLQ- 117
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 128 KTAKEGGSLIYEKGEKAFGWPGKIGAFVSIWQINIGAMSSYLFIKYELEPVIKRAFMGL 187
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 118 -----GKRVOLGVAQVGNLIGVTVGTITASISLVANGKNCFH 157
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 188 BENTGEMVINGVYLIFVSVGILLPLSLKNLGYL-GYTSGSLTGMVFVSV---VLY 242
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 158 DKGH-----TADCTISNRYMAVFGIIQVILSQIPNFHKLSPLSIMAAVMSFTYATIGI 211
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 243 KKFQIPCPPLVDHSGVNLSPNNTLPMHYVNL---PNNSSSDVNM---MDYTHN-PA 295
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 212 GLAIAFVAGKVKGTSMGTANGVDYTAQKI-----WRSFOAVGDIAPAYAVATV 263
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 296 GL-----DENQAKGSLHDSGVYEAEHSDDKCPKYFVENSRTAYAIPIIVFAFVCHPEV 349
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 264 IEIOTLRSSPAENKMKASLVGVSTTFEFYILCCIGIAAGNNAPOGFLDPGFEE 323
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 350 LPIYSELKQ-RSRRMQVTSNISTIGMLVMYLLALFGILTYGVEDELL---HAYSK 404
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 324 FWLIDFANACIAVHLIGAYOVFAQPIFOFEVEKKCNRYPNKFTSESVNVPLGKFN 383
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 405 VYTLDPILMWRLLAVAV-TLTPVILFPRI-----TSVITLLFP---KRPF 448
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 384 SLFR-LVWETAVVYITTVAMIFPFENAILGLIGAASFPLTVYFVEMHIAQTKIKYS 442
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 449 SWIRHPLIAAVILALNNVILVPTIKYIFGFIGASSATWILFDLPAVFYKLKVKETFR 508
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 443 ARMIALKTCVCLIVSLAAGSINAGLISVVTYKPEFRTH 484
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 509 S---PQKVALIFLVVGIFPMGSMALIIIDWIYDPNSKH 546
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 10
US-09-776-705-5
; Sequence 5, Application US/09776705
; Patent No. US20020082191A1
; GENERAL INFORMATION:
; APPLICANT: GUEGLER, Karl et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001010
; CURRENT APPLICATION NUMBER: US/09/776,705
; CURRENT FILING DATE: 2001-02-06
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 506
; TYPE: PRT
```

```

; ORGANISM: Human
US-09-776-705-5

Query Match      4.5%; Score 112.5; DB 10; Length 506;
Best Local Similarity 18.8%; Pred. No. 0.01;
Matches 99; Conservative 75; Mismatches 203; Indels 151; Gaps 20;

QY 2 KSFNTEGHNHSTAEGDAYVSDPTKNVDEDEKREKRTGTLTASAHITAVIGSGVLSLA 61
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 60 KKYETEFHPGTTSPGMSVNLN-----AIVSGSILGS 93
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 62 VALAQLGWIAGTSLILF-SFTYFTSTMLADCYRAPDPVTKNNYTMQVRSYIGGR 119
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 94 YAMANTG-TALPFIILTFPSIFSLYSVHLLLTANEG-----SLYLEQLGYK 140
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 120 KYQLCG-VAQVGNL-----IGTVGVTITASISLVAN-----GKNCFHDKGHADCT 166
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 141 AAGLVGKLAAGSITWQINIGAMSSYLFIKYEPLVYQALTNIEDTKGLMYLNG----- 194
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 167 ISNRYMAVFGIIQVILSQIPNFHKLSPLSIMAAVMSFTYATIGI-----GLAIATV 218
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 195 --NYLVLVSLVYLPLSLFRNLGYIGYTSGLSLCMVFLLIVICKRQVPCPVEAALI 252
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 219 AGKRVGKTSMTGT---AVGVDTA-----AQIKRSFOAVGDIAPAYAVATV 262
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 253 INETINTLTPTALVPALSHNVTTENDSCRPHYFIENSQTV---AVPILIFSFCHPA 308
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 263 LIEIOTLRSSPAENKMKASLVGVSTTFP-----YILCCIGIAAGNNAPOGFLTD 317
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 309 VLPIYEL-----KDRSRRMNNVSKISFPAFLMYLLALFGILTYEHVESLTLT 361
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 318 FGFH--EPFWLIDFANACIAVHLIGAYOVFAQPIFOFEVEKKCNRYPNKFTSESVN 374
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 362 YSILGTDILLILVRLAVLMAVTLTPVVI-----PIRSSYTHLLCAS 405
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 375 VPPLGKFNISLFRVWRYAVVI-----TTVYAMIFPPFNAILGLIGAASFPLTVYFPV 429
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 406 KDFSW-----WRHSLITVSLIAPLNLVIEPPTIRDFIGFGASAASMLFIPLPS 455
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 430 EMHIAQTK-----IKKYSARMIALKTCVCLIVSLAAGSINAGLI 471
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
DB 456 APYIKLVKPEPKMSVOKIGALFF-----LLSGVLWMTGSMALIV 494
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

RESULT 11
US-09-741-153-2
; Sequence 2, Application US/09741153
; Patent No. US20020102637A1
; GENERAL INFORMATION:
; APPLICANT: VAN, Chunhua et al
; TITLE OF INVENTION: ISOLATED HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN TRANSPORTER PROTEINS,
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: CL001059
; CURRENT APPLICATION NUMBER: US/09/741,153
; CURRENT FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Human
US-09-741-153-2

Query Match      4.4%; Score 111.5; DB 10; Length 619;
Best Local Similarity 20.5%; Pred. No. 0.017;
Matches 98; Conservative 64; Mismatches 174; Indels 143; Gaps 22;

QY 13 TASEGDATYSDPTKNVDEDEKREKRTGTLTASAHITAVIGSGVLSLMAIAQLGWI 72
   |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 18 TLSEGALET-----RLARCLSTLDLVALGVSTLGAAGVYLAGEVAKDK--AG 63
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
QY 73 TSILIFSPITFTSTMLADCYR---APDPVTKGRN-YTMDV-----VRSY- 115
   :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
```

```
Db 64 PSIVICF-LVAALSSVLAGLCYAEFGARVPCSGSAVLYSVTVVGGELWAFITGNNLILSYV 122
QY 116 LGRKVKOLCGVAGYGNLIGVTVYTTASISLVAVGKSNCFHDKGHTADCTTISNYPYMAV 175
Db 123 IGTSVARAWSSAFNDLHNIHISKTLQGGSTAL-----HVPH--- 158
QY 176 FGIIOVILSQIPNFHKLSPFLSINAAVMSP-----TVATIGIGLAIATVAGGKVGKTS 227
Db 159 -----VLAEBPFFALGLVLLGLTGLALGASESALVTKVFTGVNLLV-----LGFVM 205
QY 228 MTGTVAGVDVTAQAQKWRFSQAVGDIAFAYAYATVILIEIQDILRSSPAENKAMKRASLVG 287
Db 206 ISGFVKG-DV-----HNKLTEDYELAMA-----ELNDTVSLGLSGGVPVPGFEG 252
QY 288 V--STTTFYILCGCTGYAAGFNAPGDFLTDFGFFEPFWLIDFANAC-IAVHLIGAYQV 344
Db 253 ILRGAATCFYAFVGFDCIATTGGEAQNPORS-----IPMGIVISLVCFLAVFAVSSALT 307
QY 345 FAQPIQFQVEKKCNRNYPNKK-----FITSEYSVAVPFLGKFNIS----- 384
Db 308 LMPYIQL-----QPESPLPEAEFLYICWAPARVVAVVAGSLCALUSTSLGSMFPMRPV 359
QY 385 -----LFRLVWR-----TAYVVITTVVAMIFPPFNAIL-----GLIGAASFWPLTVY 426
Db 360 IYMAEDGLLFRVLARIHTCTRTPIATVVGIIAAMAFPLFKLTDLVLMISIGTLLAY 418
```

```
RESULT 12
US-09-815-242-13429
; Sequence 13429, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; PROKARYOTES
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 13429
; LENGTH: 440
; TYPE: PRT
; ORGANISM: Streptococcus pneumoniae
US-09-815-242-13429

Query Match 4.2%; Score 107; DB 10; Length 440;
Best Local Similarity 20.3%; Pred. No. 0.028;
Matches 88; Conservative 69; Mismatches 174; Indels 102; Gaps 21;

QY 49 ITAVIGSG-VLSLAWIAQLG-----WTAGTSILLIFSFTYFTSTMLADCYRAPDP- 99
```

```
Db 67 LASTVGTGNIIGVATAIKVGGFGALFMMWMAA-----FFGMATKYAEGLLAIKYRTKDDH 121
QY 100 -----VTGKNRYTMQVVRSY-----LGRKVKOLCGVAGYGNLIGVT--VGYTTTASIS 146
Db 122 GAVAGGPMHYILLGMGEKWRPLADLFAVAGVLVALLGIGCTFTQVNSITESIQNTTTTISPA 181
QY 147 LVAVGKSNCFHDKGHTADCTTISNYPYMAVFG---IIQVILSQIPNFHKLSPFLSINAAVM 202
Db 182 ITAL-----VLTVFVAIAVFGGLKSIKSVSTTVWP-----FMALIYYILG 220
QY 203 SPTYATIGIGLAIATVAGG-----KVGKTSMTGTAVGVDVTAQAQKWRFSQAVGDI 253
Db 221 TLTVTIFENIGKIPGTIALGFTSAFGLAAGVGAGFAGASVMAIQNGVARGVPNSGLGSA 280
QY 254 APAYAYATVLIBIQDILRSSPAENKAMKRASLVGSTTTTFFYILCGCTGY-----AAGFN 308
Db 281 PIAAAAAKT-----NEPVEQGLI---SMTGTFIDTL--IICLTGLTLTLVTVGVMSG 326
QY 309 NAPGDFLTDGFFEPFWLIDFANACIAVHLIGAYQVFAQPIQFQVEKKCNRNYPNKKPIT 368
Db 327 DLNGVALTQSAFSTVF--SHFGPALTLTFLV-----LFAFTTILGWRIYGERCF-- 373
QY 369 SEYSVNVVPLGKFNISLFRLVWRVTAYVVITTVVAMIFPPFNAILGLIGAASFWPLTVYPP 428
Db 374 -EFLGVRFIWLYRVVFLMVLLGCFIEL-NMVWIIADIVNALMAL---PNLIALLVMSP 428
QY 429 VEMHIAQTKIKKY 441
Db 429 VV--IAET--KKY 437
```

```
RESULT 13
US-09-981-353-16
; Sequence 16, Application US/09981353
; Patent No. US20020160382A1
; GENERAL INFORMATION:
; APPLICANT: Lasek, Amy W.
; APPLICANT: Jones, David A.
; TITLE OF INVENTION: GENES EXPRESSED IN COLON CANCER
; FILE REFERENCE: PA-0038 US
; CURRENT APPLICATION NUMBER: US/09/981,353
; CURRENT FILING DATE: 2001-10-11
; NUMBER OF SEQ ID NOS: 194
; SOFTWARE: PERL Program
; SEQ ID NO 16
; LENGTH: 1212
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Incyte ID No. US20020160382A1 1285632CD1
US-09-981-353-16

Query Match 4.1%; Score 104; DB 9; Length 1212;
Best Local Similarity 17.5%; Pred. No. 0.2; Mismatches 233; Indels 194; Gaps 27;
Matches 109; Conservative 87;

QY 6 TEGHN-----HSTAESGD-----AYTVSDPTKNVDEDEGREKRTGTWLTASA 46
Db 212 TFGHNTMDAVPRIDHYRHTAAQLGKLLRPSLAELHDELEKEPFEDGFANGEESTPTDA 271
QY 47 HIITAVIGSGVLSLAW-----AIAQLGWIA-----GTSILLIFSFTYF 85
Db 272 VVITYAESKGVKFGWIKGVLRVCMNLINWGMVLFRLSWIVGQAGIGLSLVIMMATVVT 331
QY 86 TSTMLADCYRAPDP--VTGKNRYTMQVVRSYLGGKVKQLCGVAGYGNLIGVTVGTITAS 144
Db 332 TITGLSTSAIATNGPVRGGAY-----YLISRLG-----PEFGAIGLITAFANAVA 379
QY 145 ISLVAVGKSNCFHDKGHTADCTTISNYPYMAVFGIIQVILSQIPNFHKLSPFLSINAAVM 202
Db 380 VAMVVVGAETVVELKEHSI-LMTDEINDIRIICAITVVI-----LLGISVAGM 428
```

Qy 203 SEFYATIGGLAIVAGK--VG-----KISMTGTAVGVDTAAQIKW 244
Db 429 EWEAKAQIVLLVLLAIDCFVIGFTIPLESKKPKGFGYKSEI FENEFDPDREBETFP 488
Qy 245 RSQNGVDIAFNAVATVLEIOTDLRSSPAENKAMKRAKSLVGVSTTFFFYI-----LCG 259
Db 489 SYF-----AIFPPATGILAGANISGDLADPGSALPKGTLLALITLTVVIAVSVGS 542
Qy 300 CIGYAFNG-----NAPG-----DFFLT-----DFGEFPEMLIDPAN----- 341
Db 543 CYVRDRTGVNDITVELNCTSAACKLNFDSSCPSSYGLMNFQVMSVSGFTPL 602
Qy 332 -----ACTAVHLIGAYOVF-----AQPFOVEKKCNRYPDNK----- 365
Db 603 ISAGIFSATLSSALSVSPKIFQALCKDNVPAFQMFAGKGVGNNEPLRGYILFPLA 662
Qy 366 ---FITSESVANVPLGKNISLFLV-----WRTAVVITTVVAMIPPF 408
Db 663 LGFILLAEINVAPIISNPFASVALINFSVFHASLAKSPGMRPG-----FKYV 711
Qy 409 NALIGLGA-----ASFWE--LTVPFVEMHIAQTKIKKYSARW-IALKTMGVCL 456
Db 712 NMWISLGLALCCTVAFVIMWMAALITVIVLGLIYVT-YKAPDVNMGSTQALTY--- 767
Qy 457 IVSLAAGSIAGLISSVKTYKP 479
Db 768 -INALQHSIRLSGVEDHVKNFRP 769

RESULT 14
US-09-895-913A-268
; Sequence 268, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1 Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; PRIOR FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/881,227
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 268
; LENGTH: 551
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-895-913A-268

Query Match 4.0%; Score 101.5; DB 9; Length 551;
Best Local Similarity 18.9%; Pred. No. 0.12; Indels 175; Gaps 23;
Matches 98; Conservative 75; Mismatches 170;

Qy 20 YTVSDPTKVNDEGRKRTGTWLTASNHITTAIVGSGVLSLAWAIAQT-GMIAGTSLILI 78
Db 5 HQYDPLGN-----WLSA-----LVALLPILFLFLSLMVFRLKQYTA----- 42
Qy 79 FSTTYFTSTMLAD-CYRAPDPVTGKKNRYMDVNSYLGGRVQLCGVAYGNLGV-T 136
Db 43 -AFLSVALSAVIAVLVYKMFVSVGS-----SFL-----YGFVLGLMP 79
Qy 137 VGYTITASISLVAVGKSNCHDGHADCTISNYPYAAVGGIIQVILISQIPNHKLSFLS 196
Db 80 IAMITIAAIFLYKLSYK-----SGT-----FELIKESYQSITLDRILIVIL 120
Qy 197 IMAAVMSFYATIGGLAIVAGKVG-----KISMTGTAVGVDTAA 240

Db 121 IGFCFSGFLEGAIGFGPIAITAAILVGLGLPYSAGLCLANTAPVAFGAGGIPISA- 179
Qy 241 QKIMRSFOAVVDIAFVAVATVLEIOTDLRSSPAENKAMKRAKSLVGVSTTFFFYILCCG 300
Db 180 -----MASAVGPAILI-----SMTGKILFVSLVDPFFIVFLM 214
Qy 301 IGYAFGNNAPODFLTDFGFEFPEMLID-----PANACTAVHLIGAYOVAOP-I 349
Db 215 DFKKIKETFPVAFVIAFSPAGQFLSNVYLGPELFGIISALVSVATLFLKFWQPKAI 274
Qy 350 PQFVEKKCNRYPDNKFTITSESVANVPLGKNISLFLVWRTAVVITTVVAMIPPF 409
Db 275 FSDGKAAS-----FTKSNH-----HICKIVAVMSPFVILVIVLMIOPFFK 317
Qy 410 ALL-----GLGAASF-----PL-----TYFPEVEMHIAQTKIKKYSARWIA 447
Db 318 ALPEKDGILAFSNFYFEPNNISNHIFKSPFVEANOSVSPVVF-----KFL 365
Qy 448 LKTM--CYVCLIVSLAAGSIAGLISSV-KTYKPR 481
Db 366 INTVGSIRLALVSMVLVRAVSDALSVFGELKMR 403

RESULT 15
US-09-817-514A-2
; Sequence 2, Application US/09817514A
; Patent No. US20020078478A1
; GENERAL INFORMATION:
; APPLICANT: French-Constant, Richard
; APPLICANT: Bowen, David
; APPLICANT: Rochelleau, Thomas
; APPLICANT: Waterfield, Nicholas
; TITLE OF INVENTION: DNA SEQUENCES FROM PHOTORHABDUS LUMINESCENS
; FILE REFERENCE: 61645
; CURRENT APPLICATION NUMBER: US/09/817,514A
; PRIOR FILING DATE: 2000-03-26
; PRIOR APPLICATION NUMBER: US 60/191806
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2
; LENGTH: 2516
; TYPE: PRT
; ORGANISM: Photorhabdus luminescens
US-09-817-514A-2

Query Match 3.8%; Score 96; DB 10; Length 2516;
Best Local Similarity 25.1%; Pred. No. 3;
Matches 50; Conservative 32; Mismatches 67; Indels 50; Gaps 10;

Qy 238 TAAQKIMRSFOAVVDIAFVAVATV--LEIOD---TRSSPAENKAMKRAKSLVGVSTT 291
Db 183 TKVEMLSFTRPSGATPYHADVENREVIOQDPLEQUNASPAIAGLWQASLGLGINS 242
Qy 292 ---TFFYILCGCIGYAFGNNAPODFLTDFGFEFPEMLIDPANACTAVHLIGAYOVAOP 348
Db 243 ISPELFLNLTETI-----TGNABEELYKKNFGNIEP-----ASLAMPETLKRYINISDEE 292
Qy 349 IFQFVEKKCN--RNPDPKF-----ITSESVNV-----PFLGKR 361
Db 293 LSGFIKASNPFGQSESNNOQLITPVNUSDGTVKYVRIITREYTNAYQMDVELFPFGGE- 351
Qy 382 NISL---FLVWRTAVVI 397
Db 352 NYRLDYKRFKNFYASYSI 370

Search completed: November 22, 2002, 13:32:10
Job time : 15 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2002, 13:26:12 ; Search time 14 Seconds

(without alignments)
1436.859 Million cell updates/sec

Title: US-09-854-562-2

Perfect score: 2519
Sequence: 1 MKSPNTEGHNHSTAESGDY.....SIAGLISSVKTYPKPTNHE 485

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	208	8.3	713	YMK1_YEAST	P50944 saccharomyc
2	207.5	8.2	470	MTR_NEUCR	P38680 neurospora
3	189.5	7.5	692	YX06_YEAST	P36622 saccharomyc
4	167.5	6.6	602	YX01_YEAST	P47082 saccharomyc
5	165.5	6.6	486	YMA47_CAEEL	P34479 caenorhabdi
6	145	5.8	460	YMA32_CAEEL	P34479 caenorhabdi
7	142	5.6	656	YAN9_SCHPO	Q10074 schizosacch
8	131	5.2	490	YI18_YEAST	P40501 saccharomyc
9	128.5	5.1	565	YD08_MYCPN	P75472 mycoplasma
10	125.5	5.0	480	YEH4_YEAST	P39981 saccharomyc
11	121.5	4.8	639	FU11_YEAST	P38196 saccharomyc
12	121	4.8	440	YFPA_SCHPO	Q9018 schizosacch
13	120	4.8	440	YU99_MYCTU	Q10858 mycobacteri
14	118.5	4.7	516	YWCA_BACSU	P39599 bacillus su
15	115.5	4.6	615	YSPK_CAEEL	Q19425 caenorhabdi
16	114.5	4.5	422	EXUT_BACSU	Q34456 bacillus su
17	114	4.5	1205	NKCI_MOUSE	P55012 mus musculu
18	108	4.3	1212	NKCI_HUMAN	P55011 homo sapien
19	107	4.2	1023	TSCC_PSEAM	P55019 pseudopieur
20	106.5	4.2	490	YWOE_BACSU	P94429 haemophilus
21	105.5	4.2	388	HCA7_HAEIN	P55014 mus musculu
22	105.5	4.2	1095	NKCA_MOUSE	P30818 clostridium
23	105	4.2	481	Y389_CLOPE	Q92897 rickettsia
24	105	4.2	657	NUOL_RICCN	P42086 bacillus su
25	104.5	4.1	438	PBUX_BACSU	Q12300 saccharomyc
26	104	4.1	763	YBAT_YEAST	P55010 escherichia
27	102.5	4.1	430	YBAT_ECOLI	P55010 escherichia
28	102.5	4.1	1095	NKCA_RAT	P55010 escherichia
29	102	4.0	705	NUSC_HORVU	Q12340 hordeum vul
30	101.5	4.0	448	GNTP_BACLI	P46832 bacillus li
31	101.5	4.0	580	ISPS_SCHPO	P40901 schizosacch
32	101.5	4.0	633	FUR4_YEAST	P05316 saccharomyc
33	101	4.0	465	ALST_BACSU	Q45068 bacillus su

34	100.5	4.0	488	1	SECY_HALVO	Q977V3 halobacteri
35	100.5	4.0	702	1	NU5C_SORBI	Q33066 sorghum bic
36	100.5	4.0	738	1	NU5C_MAIZE	P46620 zea mays (m
37	100.5	4.0	1191	1	NKCI_SOUAC	P55013 squalus aca
38	99.5	3.9	407	1	RFC_SALTU	P26479 salmoneilla
39	99	3.9	608	1	DSBD_RALSO	Q8XV41 raietonia s
40	99	3.9	744	1	NU5C_GERBA	P51100 gerbera jam
41	99	3.9	809	1	CICH_TORCA	P35520 torpeda cal
42	98.5	3.9	426	1	Y680_CHLPN	Q9Z7M4 chlamydia p
43	98.5	3.9	1581	1	VGLP_BEV	P23052 berne virus
44	98	3.9	440	1	YHUE_ECOLI	P37643 escherichia
45	98	3.9	608	1	DIPS_YEAST	P53388 saccharomyc

ALIGNMENTS

RESULT 1
YMK1_YEAST STANDARD; PRT; 713 AA.
ID AC P50944;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 80.0 kDa protein in POLI-RAS2 intergenic region.
GN YNL101W OR N2185.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RX MEDLINE=96267765; PubMed=8701612;
RA Saiz U.B., Buitrago M.U., Soler A., del Rey F., Renault J.L.;
RT "The sequence of a 21.3 kb DNA fragment from the left arm of yeast
RT chromosome XIV reveals LEU4, MET4, POL1, RAS2, and six new open
RT reading frames."
RL Yeast 12:403-409(1996).
CC - SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC - SIMILARITY: TO YEAST YKL146W AND S.POMBE SPAC3H1.09C.
CC -
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: Z50161; CAA90525.1; -;
CC EMBL: Z71377; CAA95977.1; -;
CC SCD: S0005045; YNL101W.
CC InterPro: IPR002422; AA/re1_primease2.
CC Pfam: PF01490; Aa_trans_1.
CC Hypothetical protein; Transmembrane.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 327 347 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 411 431 POTENTIAL.
FT TRANSMEM 439 459 POTENTIAL.
FT TRANSMEM 484 504 POTENTIAL.
FT TRANSMEM 516 536 POTENTIAL.
FT TRANSMEM 562 582 POTENTIAL.
FT TRANSMEM 622 642 POTENTIAL.
FT TRANSMEM 649 669 POTENTIAL.
FT TRANSMEM 693 713 POTENTIAL.
SQ SEQUENCE 713 AA; 80025 MW; EAH458A2FEB660FB CRC64;
Query Match 8.3%; Score 208; DB 1; Length 713;
Best Local Similarity 21.4%; Pred. No. 1.8e-07;
Matches 118; Conservative 75; Mismatches 197; Indels 162; Gaps 22;

```

QY 5 NTEGHNHSTAESGDATVSDPTKQVDED----- 32
DB 207 NDSASDFTSHESDSINQSSPSSNQDIDKVPFLTRNPLEFLYVGFHAFESFEDDIPDS 266
QY 33 -----GREKRT-----GTWLTASAH11--TAVIGSGVLSLAWAIAQLG 68
DB 267 SNMIRGEDESRALLSRPDHMKVLPKAGTSTTKVFLILLKSFITGVFLFNAPFHNGG 326
QY 69 WIAGTSSILLIFSPITYFTSTMLADCYRAPDPVTKGRNYTMDVVRYSYLGGRKVQLCGVAQ 128
DB 327 LFFSVSLAFFGIYSYV-----CY-----YILVOAKSCGVSS 359
QY 129 YGN-----LIGTVGYTITASIVAVGKS-----NCFHDKGTADCTIS 168
DB 360 FGDIGLKLYGPMWR111LFSLVITQVFGSGAYMIFTAKNLOAFLDNVFF-----VGLV 412
QY 169 NYPYMAVF-GIIQVILSQIPNFHKLFLS1MAAVMSFTVATIGIGLAIATVAGKVGKTS 227
DB 413 PLSYLMVFQTIPIPLSFTRNISKLSLPSLLANFF-----IWAGLVIVIIPTAKELFPD 466
QY 228 MTGT-AVGVDVTAAQKIWSFQAVGDIAPAYATVLLIEIQDTRLRSPAENKAMKRAASLV 286
DB 467 LMGTPAMGVVYGLNADRWTLF--IGTAIFAEGIGLIIIPQDSMRN---PEKFPPLVALV 521
QY 287 GVSTTTFFVILCGICGYAAGFNAPGDFLTDGFPFPPFLIDFANACIAVHLIGAY----- 342
DB 522 -ILTATILFISATLGYLAGSVN-----QTVILLNLPOSNIFVNLQLFYSTA 569
QY 343 -----QVPAQIFOFVEK-----KCNRYPNKFTITSEYVNVNPFGLKFNISLRLVNR 391
DB 570 IMLSTPLQLF--PAKLIENKFPFKTKIYVKHDDITRVELR-PNSGKLN---WKIKWL 623
QY 392 TAYV--VITTVAMIFPP-----FNAILGJIGAASFVPLTVYFVPEMHIAQTKIKKYSARW 445
DB 624 KNFIRSIIVVSVIAYFGSDNLDKFSVIGSLACIPLVVIYPSMLHRLGNSLPETKGEF 683
QY 446 IALKTWYCVCLI 457
DB 684 WRPKMLDTILI 695

```

RESULT 2

```

MTR_NEUCR STANDARD; PRT; 470 AA.
ID MTR_NEUCR AC P38680;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE N amino acid transport system protein (Methyltryptophan resistance protein).
GN MTR.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Oak Ridge;
RX MEDLINE=95095055; PubMed=8001794;
RA Dillon D., Stadler D.;
RT "Spontaneous mutation at the mtr locus in neurospora: the molecular spectrum in wild-type and a mutator strain."
RL Genetics 138:61-74 (1994).
RN [2]
RP SEQUENCE OF 210-470 FROM N.A.
RX MEDLINE=92146948; PubMed=1838345;
RA Koo K., Stuart W.D.;
RT "Sequence and structure of mtr, an amino acid transport gene of Neurospora crassa."
RL Genome 34:644-651 (1991).
CC -!- FUNCTION: REQUIRED FOR THE TRANSPORT OF NEUTRAL ALIPHATIC AND AROMATIC AMINO ACIDS VIA THE N SYSTEM.

```

```

CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC EMBL; L34605; AAA33600.1; -.
CC EMBL; S81767; AAB21410.1; -.
CC PIR; A54551; A54551.
CC InterPro; IPR002422; AA/rel_pmasease2.
CC Pfam; PF01490; Aa_trans; 1.
CC KW Transport; Amino-acid transport; Transmembrane.
CC FT DOMAIN 1 56 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 57 77 POTENTIAL.
CC FT TRANSMEM 78 98 POTENTIAL.
CC FT DOMAIN 99 131 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 132 152 POTENTIAL.
CC FT DOMAIN 153 168 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 169 189 POTENTIAL.
CC FT TRANSMEM 191 211 POTENTIAL.
CC FT DOMAIN 212 236 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 237 257 POTENTIAL.
CC FT DOMAIN 258 275 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 276 296 POTENTIAL.
CC FT DOMAIN 297 316 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 317 337 POTENTIAL.
CC FT DOMAIN 338 357 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 358 378 POTENTIAL.
CC FT DOMAIN 379 386 CYTOPLASMIC (POTENTIAL).
CC FT TRANSMEM 387 407 POTENTIAL.
CC FT DOMAIN 408 427 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 428 448 POTENTIAL.
CC FT DOMAIN 449 470 CYTOPLASMIC (POTENTIAL).
CC SQ SEQUENCE 470 AA; 51162 MW; E8132D1A62373300 CRC64;

```

```

Query Match 8.2%; Score 207.5; DB 1; Length 470;
Best Local Similarity 22.3%; Pred. No. 1.2e-07;
Matches 110; Conservative 75; Mismatches 192; Indels 117; Gaps 24;

QY 31 EDGREK--RTGTLWTASAHIIITAVIGSGVLSLAWAIAQLGWIAGTSSILLIFSYFTST 88
DB 44 QEGHAKPHRLG-WKRLTVVLIVEAIALGSLSPGAFATLGMVPGVILSVGMGLICIYTAH 102
QY 89 MLADCYRAPDPVTKGRNYTMDVVRYSYLGGRKVQLCGVAQVGNLIGVTGVYTTITASISLV 148
DB 103 VIGQTKLKHPEIA-----HYADVGRVMEGRGVEIISFMVLQLLFIYVGSVHLTGIMWG 157
QY 149 AVGKSNCFHDKGTADCTISNYPYMAVFGIIQVI-----LSQIPNFHK-----LSFLSIM 198
DB 158 TI-----TDNGN-GTCSL-----VFGIVSAIILFLLAIPPSFAEVAAILGYIDFVSIC 203
QY 199 AAVMSFTVATIGIGLAIATVAGKVGKTSMTGTAIV-----DVTAAQKIWSFQAV 250
DB 204 AAIL-----ITMIATG-IRSSHQEGGLAAVPMSCWPKEDSLAE-----GFTAV 246
QY 251 GDIAFAYATVLLIEIQDTRLRSPAENKAMKRAASLVGVSTTTFFVILCGICGYAAGFNA 310
DB 247 SNIVFAYFAMQCFSDMBEMHTPSDYKKSIVVALGLIEI-----FIYTVGGVYAFVGEV 302
QY 311 PGDFLTD-----FGFEPFPLIDFANACIAVHLIGAYQVFAQPIQFQVEKCKNRY 361
DB 303 QSPALLSAGPLAKVAFGLALP---VIFISGINTVVSRYLI-----ERIW 346
QY 362 PNKFIITSEYVNV-----FLGKNISLRLVWRTAVTITTVVAMIFPPFNAILGLI- 415
DB 347 PNN---VIRY-VNTPAGMWVWLGP-PDFGITLIW-----VIAEAIPPFSDLAICS 392
QY 416 -----GAASFVPLTVYFVPEMHIAQTKIKKYSARWIAKTMICYV---CLIVSLAAAGS 466

```

```

Db 393 ALFISGFSFFPALMFKITRNDASQCKKFLD--ALNMLCFVIGMGLIGTYAIAID 450
QY 467 IAGLISSVKTYPF 480
DB 451 IMDRYDHGKVKSPY 464

RESULT 3
YX06_YEAST STANDARD: PRT: 692 AA.
AC P36062:
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Hypothetical 75.5 kDa protein in SDH1-CIM5/YTA3 intergenic region.
GN YKJ146W OR YKJ600.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C;
RX MEDLINE=94378720; PubMed=8091859;
RA Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Higer F.;
RT "DNA sequencing of a 36.2 kb fragment located between the FAS1 and
RL LAP loci of chromosome XI of Saccharomyces cerevisiae.";
RN Yeast 10:835-840(1994).

[2]
RP SEQUENCE OF 447-692 FROM N.A.
RA Rad M.R., Xu G., Kirchrath L., Filtz C., Keuchel H., Hollenberg C.P.;
RL Submitted (MAR-1994) to the EMBL/Genbank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO YEAST YNL101W AND S.POMBE SPAC3H1.09C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z26877; CAA81508.1; -
DR EMBL: Z28146; CAA81988.1; -
DR PIR: S37976; S37976.
DR SGD: S0001629; YKJ146W.
DR InterPro: IPR002422; AA/rel_primease2.
DR Pfam: PF01490; Aa.trans; 1.
DR KW Hypothetical protein; Transmembrane.
FT DOMAIN 261 276 POLY-GLU.
FT TRANSMEM 302 322 POTENTIAL.
FT TRANSMEM 329 349 POTENTIAL.
FT TRANSMEM 374 394 POTENTIAL.
FT TRANSMEM 412 432 POTENTIAL.
FT TRANSMEM 443 463 POTENTIAL.
FT TRANSMEM 483 503 POTENTIAL.
FT TRANSMEM 519 539 POTENTIAL.
FT TRANSMEM 561 581 POTENTIAL.
FT TRANSMEM 607 627 POTENTIAL.
FT TRANSMEM 630 650 POTENTIAL.
FT TRANSMEM 665 685 POTENTIAL.
SO SEQUENCE 692 AA; 75459 MW; 379054069094A0F5 CRC64;

Query Match 7.5%; Score 189.5; DB 1; Length 692;
Best Local Similarity 23.1%; Pred. No. 3.5e-06;
Matches 106; Conservative 73; Mismatches 183; Indels 97; Gaps 18;
QY 8 GHHSHTAESDQATVDSPTKVV-----DEDEKERTGTWLTSAHITTAIVGSGV 57
DB 256 GBDLSEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESEEESE 315
QY 58 ISLAWAIQLQGLWIGTASILLIFSPITY--FTSTWLDQCYRAPDPVTGKRNVTYMDVRSY 115

```

```

Db 316 LFLPRAFNHNGWGFSAALCLLSALISYGFWSLI-----TTXKRV-- 355
QY 116 LGGRKVLQGVAYQGN-----LIGTVGTYITASISLVAVGKS-----NC 155
DB 356 -----GVDGYDMGRILYGPCKRKALISLISQIGSBAATVFTATNLQVSENF 406
QY 156 FHDKGTADCTISNYPYNAVFGIIQVILISQIPNFKLSFLSMAVMSFTYATIGLAI 215
DB 407 FLTK--PGSISLATYIFAQV--LIFVPLSLFRNIAKLSTGAILADLF-----ILGLVY 456
QY 216 ATVAGKRGKTSMTGTANGVDVTAQAQKIRSRQAVGDIAPFAVATVLIHIDOTLRSSA 275
DB 457 VVVY--SIYVIAVNSVADPTMFMFKADWSLF--IGTAIFTEGIGLIPIQESMK---- 508
QY 276 ENKAKKRAISLVGV-STTFEFTYLLGCGICGYAAGNNAAPGDLTDFGFEFPWMLIDPANA 334
DB 509 -HPKXFRSLAVMCLIVAVIFISGLLCTAAGSDVKVTLINFPQDYSYTLTVQLLVA 567
QY 335 AVHLIGAYQVFAQPIFOFVEKKCNRNYPDNKEITSEYSVNPFLGKFNISLRLVWRAY 394
DB 568 ALLSTPLQLF--PAIRILE--NMTPPSN--ASQKYNPKVKMLNY-----FRCAI 612
QY 395 VITTVVAMI-FPFNAILGLIGAASFWEPLTYFPEVEMH 432
DB 613 VVLTSLIAVGVANDDKFVSLVGSFACIPLIYVPLLH 651

RESULT 4
YX01_YEAST STANDARD: PRT: 602 AA.
AC P47082;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Hypothetical 65.3 kDa protein in PR3-MEP10 intergenic region.
GN YJ001W OR J1409 OR YJ083.4.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / FY1679;
RA de Haan M., Smits P.H.M., Grievell L.A.;
RL Submitted (MAY-1995) to the EMBL/Genbank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO C.ELEGANS UNC-47.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X87611; CAA60922.1; -
DR EMBL: Z49501; CAA89523.1; -
DR SGD: S0003761; YJ001W.
DR InterPro: IPR002422; AA/rel_primease2.
DR Pfam: PF01490; Aa.trans; 1.
DR KW Hypothetical protein; Transmembrane.
FT TRANSMEM 220 240 POTENTIAL.
FT TRANSMEM 241 261 POTENTIAL.
FT TRANSMEM 287 307 POTENTIAL.
FT TRANSMEM 322 342 POTENTIAL.
FT TRANSMEM 345 365 POTENTIAL.
FT TRANSMEM 390 410 POTENTIAL.
FT TRANSMEM 430 450 POTENTIAL.
FT TRANSMEM 467 487 POTENTIAL.
FT TRANSMEM 518 538 POTENTIAL.
FT TRANSMEM 544 564 POTENTIAL.

```

```
FT TRANSMEM 581 601 POTENTIAL.
SQ SEQUENCE 602 AA; 55345 MW; C2F321E76C21C2B4 CRC64;

Query Match 6.6%; Score 167.5; DB 1; Length 602;
Best Local Similarity 19.6%; Pred. No. 0.00011;
Matches 99; Conservative 92; Mismatches 208; Indels 105; Gaps 19;

QY 17 GDAYTVSDPTKNVDE-----DREKRTGTWLTASA-----46
DB 156 GDSGSLRPTASLTEMWGGAGRRFTNNDMSIVKRVEGVGDKVVTLLAGOSTAPQTIF 215

QY 47 HITAVIGSVLSLAWAIAQLGWIAGTSILLIPSFITYFTSTMLADCYRAPDPVTGKRVY 106
DB 216 NSINVLIGLLALPLGLKAGVHIGLWLAIFALATFCTAELLSCLOT-DPTL-----I 270

QY 107 TYMDVRSYLGGRKVLGCGVAQGNLIGVTVGTITASISLVAVGSKNCFHDKGHTADCT 166
DB 271 SYADLGYAAPGTKRALISALFTLDLLGSGVSLVILFGDSLNAL-----314

QY 167 ISNYPYMAV--FGIIQVILSOINPFLKLSPLS--IMAAVMSFTYATIGIGLATATVAGGK 222
DB 315 ---FPOYSTTFKIVSFFVTPVPFPLSVLSNISLLGLIS-----FTGTVLVICCCG---364

QY 223 VGKTSMTGTAVGVDDVTAQAQIWM-----RSFQAVGDIAYAYATVLEIQTDLRSPAE 277
DB 365 LYKSSSPGSLVNPMTS---MWFIDLKHLCISLISACWGCHAVFPNLTDMRHDPDKFK 421

QY 278 KAMKR-----ASLVGVSTTFYFYLCCGIGYAFGNAPGDFLDFGFFE--PFWLIDFAN 331
DB 422 DCLTKYTKITSVTDIGTAV-----IGFLMFGNLVKDEITKNVLLTEGYPKFVYGLIS 473

QY 332 ACTAVHLIGAYQVFAPIQFQVEKKNRNPDKFITSEYSVNVPFLGKFNISLFLVWR 391
DB 474 ALMTIIPKATPLNAPRPIVSLVDLMNVQHD-----EASAIRKRAAKGLQVFNRI 527

QY 392 TAVVITTVVAMIFPPFNAILGLIGAASFWPLTVYFP--VEMHIAOTKIKKYSARWIALK 449
DB 528 N---VVFVLIANFPEDKLIAPLGLAGLCTICLIILPCWFYLRKCTTKPWE-----R 578

QY 450 TMCYVCLIVS-LIAAAGSIAGLIS 472
DB 579 VACHVTICISVLSLTLGVGAALIS 602

RESULT 5
UN47 CAEBL STANDARD; PRT; 486 AA.
AC P34579; O17475;
DT 01-FEB-1994 (Rel. 28, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Uncoordinated protein 47 (protein unc-47).
GN UNC-47 OR T20G5.6.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98007977; PubMed=9349821;
RA McIntire S.L., Reimer R.J., Schuske K., Edwards R.H., Jorgensen E.M.;
RT "Identification and characterization of the vesicular GABA
transporter."
RL Nature 389:870-876 (1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Berks M., Smith A.;
RN [3]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
```

```
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: TO YEAST YJR001W.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF031935; AAB87066.1; -
DR EMBL; Z30423; CAA83006.2; -
DR PIR; S42372; S42372
DR WormPep; T20G5.6; CE25119.
DR InterPro; IPR002422; AA/rel_pmasease2.
DR Pfam; PF01490; Aa_trans; 1.
KW Transmembrane.
FT TRANSMEM 94 114 POTENTIAL.
FT TRANSMEM 120 140 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 204 224 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 264 284 POTENTIAL.
FT TRANSMEM 306 326 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 399 419 POTENTIAL.
FT TRANSMEM 422 442 POTENTIAL.
FT TRANSMEM 458 478 POTENTIAL.
SQ SEQUENCE 486 AA; 54439 MW; 6435C3B338552925 CRC64;

Query Match 6.6%; Score 165.5; DB 1; Length 486;
Best Local Similarity 19.4%; Pred. No. 0.00012;
Matches 97; Conservative 91; Mismatches 183; Indels 129; Gaps 22;

QY 13 TAESG-----DAYTVSDPTK---NVDEDGREGKRTG-----TWLTASAHIIT 50
DB 41 TGESSNPPPHDRLEPIQESVVSQPOKDDINKQBEAKDDGHGEASEPISALQAAWNVTN 100

QY 51 AVIGSVLSLAWAIAQLGWIAGTSILLIPSFITYFTSTMLADCYRAPDPVTG-KENYTYM 109
DB 101 AIOGMFIVGLPIAVKVGWMS-IGAMVGVAVCYVWTVLLIECLVE---NGVKRKTVR 155

QY 110 DVRSYLGGRKVLGCGVAQGNLIGVTVGTITASISLVAVGSKNCFH--DKGHTADCTI 167
DB 156 EIADFYKPGFGKWL-AAQLTTELLSTCIYLVLAADLL-----QSCFESVDKAGHMMITS 209

QY 168 SNPYMAVFGIIQVILSOINPFLKLSPLS-----MAAVMSFTYATIGIGLAIAITVA 219
DB 210 ASLLTCSFLDDLIQIV-SRLSPFNALSHLVNLIMVLYCLSFVSQWSFTITFSLNINTLP 268

QY 220 GGVKGTSMGTAVGVDDVTAQAQIWRSPQAVGDIAYAYATVLEIQTDLRSPAEKNA 279
DB 269 -----TIVGMVVFYGTSHIFLPLNLEGNMK-NPAQPNV 299

QY 280 MKRASLVGSVTTTFYILCGGIGYAAFG-----NNAPGDFLTDGFFFPFWLIDFAN 331
DB 300 MLKWSHIAAAV---FKVVFGLMGLFTFGELTQEETISNLPN-----QSEKIL--VN 345

QY 332 ACTAVHLIGAYQVFAPIQFQVEKKNRNPDKFITSEYSVNVPFLGKFNISLFLVWR 391
DB 346 LILVVKALLSLPLPYAAVQLKKNNLFLGYQPTFP-TSCYSPD-KSLREWAVTL-----R 398

QY 392 TAYVVITTVVAMIFPPFNAILGLIGAASFWPLTVYFPVEMHIAOTKIKKYSARWIALKTM 451
DB 399 IILVLTFLVALSVPLVELMGLVGNITGTMLSPWLPALFHL-----YIKEXTL 447

QY 452 -----CYVCL 456
DB 448 NNFEKRFQGGIIMGCSYCI 467

RESULT 6
```


YMU2_CAEEL STANDARD; PRT; 460 AA.
 ID YMU2_CAEEL 093E0;
 AC P34479; Q93E0;
 DT 01-FEB-1994 (Rel. 28, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Hypothetical protein F59B2.2 in chromosome III.
 DE F59B2.2.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OC NCBI_TaxID=6239;
 RX MEDLINE=94150718; PubMed=7906398;
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berke M.,
 Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
 Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
 Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
 Johnston L., Jones M., Kersey J., Kirsten J., Laister N.,
 Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
 Parsons J., Percy C., Rifkin L., Roopia A., Saunders D., Shownkeen R.,
 Sims M., Smailton N., Smith A., Smith M., Sonhammer E., Staden R.,
 Sulston J., Thierry-Mieg J., Thomas K., Vaughan K.,
 Waterston R., Watson A., Weinstock L., Wilkinson-Sprot J.,
 Woldman P.;
 RA "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
 elegans.";
 RL Nature 368:32-38(1994).
 RN [2]
 RP REVISIONS.
 RA Durbin R.;
 RU Submitted (DEC-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL: 211505; CAA7582.1; -
 DR PIR: S31123; S31123.
 DR WormPep: F59B2.2; CE20899.
 DR InterPro: IPR002422; AA/rel_primease2.
 DR Pfam: PF01490; Aa.trans; 1.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 64 84
 FT TRANSMEM 132 152 POTENTIAL.
 FT TRANSMEM 172 192 POTENTIAL.
 FT TRANSMEM 194 214 POTENTIAL.
 FT TRANSMEM 236 256 POTENTIAL.
 FT TRANSMEM 270 290 POTENTIAL.
 FT TRANSMEM 316 336 POTENTIAL.
 FT TRANSMEM 362 382 POTENTIAL.
 FT TRANSMEM 383 403 POTENTIAL.
 FT TRANSMEM 426 446 POTENTIAL.
 SQ SEQUENCE 460 AA; 51338 MW; 0D7809A84EF15BA CRC64;
 Query Match 5.8%; Score 145; DB 1; Length 460;
 Best Local Similarity 20.3%; Pred. No. 0.0031;
 Matches 109; Conservative 79; Mismatches 186; Indels 164; Gaps 28;
 QY 7 EGNHSTAESGAYTSDPTKNVDEGRKXTWLTASAHIIT---AVIGSVLST--A 61
 DB 5 EGRBRAVTEGDSKSNMGALVQPPA---RSGDVITPTAAVLTLSKSMNACGFSIPYA 61
 QY 62 WALAQGLWAGTSTILLIFSFTYFTSTMLADCYRAPDPVTKGNNTYMDVRSYLGKRV 121
 DB 62 WKLGGL-WVS-----FVMSFV----- 76

QY 122 QLCGVAOYGNLIGVTGTTASISLVAVGKSNCHDKGHTAD--CTISNPYM-----A 174
 DB 77 -IAGLNWYGNHILVRAQGLAKKSDRSAL-----DYGHFAKKYCDYSDIRFLNNNSKA 128
 QY 175 VEGIIQVILSQIPNPKHSFSLMAAVMSFTYATIGG-----LAIATYAGGRV 223
 DB 129 VWFYFNVITL---FYQIGMCSVALFLSDNLVNL-VSDHLGSTRHQQIMLWATVSLFPI 183
 QY 224 GKTSM-----TGTVGVDPVTAQ-KIMSPQA-----VGD1 253
 DB 184 LITNFTKRIYSFPALVSSVFVIGAAVIMQYTOQRPQMDKLAATNFTGTIMIGNS 243
 QY 254 AFAYAVATVLEIQTLLRSSPAENKMRKASLVGST---TFYIILGCGIYAFGNN 309
 DB 244 WYAFSGQWMLIPLENKL-----DNPAALAPRGVLTMTIICATPMTALGFGYTGFGDS 298
 QY 310 AGDFLITDF---GFEP---FWLIDFANACIAHLIGAYVFAQPIQFVKKCKRNP 362
 DB 299 IAPITTYNPKKGLYSTVNVFLMLQSLGNSIAMYVV--YDMF---FNGFRKKGARPP 352
 QY 363 DKKFTISEVSVNP-FLGKFNISLRLVWRTAVVITVWAMIPFENAILGLIGASFW 421
 DB 353 -----NVPKWLSDKG--FRVF---VLTVMVAVLIRLEIMPLVGVTSQA 394
 QY 422 PLTVFPP--VENHIAQTKIK--KYSARWIALKTCVYCLIVSLAAGSIALSSV 474
 DB 395 LCAALFPFFEFMTWTMTDKKGLITROR---MTKIFILVMAIGVFAIINGVITNI 448
 RESULT 7
 YAN9 SCHPO STANDARD; PRT; 656 AA.
 ID YAN9 SCHPO 010074;
 AC 010074;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 GN Hypothetical protein C3H1.09c in chromosome I.
 DE SPAC3H1.09c.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OC NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RX MEDLINE=21848401; PubMed=11859360;
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
 Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
 Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
 Collins M., Connor R., Cronin A., Davis P., Felwell T., Fraser A.,
 Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
 Holroyd S., Hornby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,
 James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
 Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
 Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitch E.,
 Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
 Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
 Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
 Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
 Weljens I., Vanstreels E., Rieger M., Schaefer M., Meller-Auer S.,
 Gabel C., Fuchs M., Fritz C., Holzer E., Meisel D., Hiltbert-Huer S.,
 Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
 Eger P., Zimmermann W., Wedler H., Wambut R., Purrelle B.,
 Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
 Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Huret S.M.,
 Lucas M., Rochet M., Gallardin C., Tallada V.A., Garçon A., Thode G.,
 Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
 Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forstburg S.L.,
 Cernutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
 Shpakovski G.V., Useery D., Barrett B.G., Nurse P.;
 RA "The genome sequence of Schizosaccharomyces pombe.";


```

Db 312 DPQVSEYTRASNLRIPISMTVEDPAQPSDALDATSYNEOCLPENGNF-DNGSIESQENN 370
Qy 375 -----VPLFGKFNISLFLVWRPAAVYIT-VVAMFPFPNAILIGIA 418
Db 371 DERGTAAVAGDNEHAPFVK-----RFRWTALLIMYTLALVSQFALVLSVGT 424
Qy 419 SFMPLTVFP 428
Db 425 GSTSISFTLP 434

RESULT 9
YD08_MYCPN STANDARD; PRT; 565 AA.
AC P75472; O08088;
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Hypothetical protein MPN308 (F10_orf565).
GN MPN308 OR MP528.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_Taxid=2104;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirk E., Li B.-C.,
RT Hermann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RL pneumoniae."
RL Nucleic Acids Res. 24:4420-4449 (1996).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97252497; PubMed=9098066;
RA Krause D.C., Proft T., Hedreya C.T., Hilbert H., Plagens H.,
RA Hermann R.;
RT "Transposon mutagenesis reinforces the correlation between Mycoplasma
RT pneumoniae cytoskeletal protein Hmw2 and cytoadherence."
RL J. Bacteriol. 179:2668-2677 (1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: IN THE N-TERMINAL SECTION, TO M.PNEUMONIAE MPN095.
CC -1- SIMILARITY: IN THE C-TERMINAL SECTION, TO M.PNEUMONIAE MPN096.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000052; AAB96176.1; -
DR EMBL; U59896; AAB52525.1; -
KM InterPro: IPR002293; AA/rel.primase.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 8
FT TRANSMEM 28
FT TRANSMEM 43
FT TRANSMEM 63
FT TRANSMEM 95
FT TRANSMEM 115
FT TRANSMEM 137
FT TRANSMEM 157
FT TRANSMEM 167
FT TRANSMEM 187
FT TRANSMEM 227
FT TRANSMEM 247
FT TRANSMEM 268
FT TRANSMEM 314
FT TRANSMEM 334
FT TRANSMEM 367
FT TRANSMEM 387
FT TRANSMEM 424
FT TRANSMEM 444
FT TRANSMEM 460
FT TRANSMEM 480
FT TRANSMEM 482
FT TRANSMEM 502
FT TRANSMEM 516
FT TRANSMEM 536
SQ SEQUENCE 565 AA; 62560 MW; ACC745B43D2184E2 CRC64;

```

```

Query Match 5.1%; Score 128.5; DB 1; Length 565;
Best Local Similarity 19.5%; Pred. No. 0.057;
Matches 99; Conservative 69; Mismatches 165; Indels 175; Gaps 23;

Qy 41 WLTAHITITANIGSGVSLAMAINOGLMIAGTSTLLIFSFITFTSTMLACYAPADPV 100
Db 48 WLVAASV-----AVTAMALALIEIASVRDN-LSTISWKEVNRRL-----Y 88
Qy 101 TGRKNY-TYMDVRSYSLGGRKQVQLGVAQ-YGNLIGTVG-----YTITASISLVAV 150
Db 69 HCGKNFMVLYLPPLFFPMPLYFICSIDGFRGLGLGTGAFHFNSTVMDLWLALALI-- 146
Qy 151 GKSNCFHDXKHTADCTISNY-----PYNAVFGIIVIIISQIPNFKLSPLSTMAAVMFT 205
Db 147 -----ITTYFPLTIPPLAKVGNIONMVSAVKFIDLVFVP----- 182
Qy 206 VATTIGLAIATVAGKV-----GKTSMTGTAAGVDVT 238
Db 183 -----IGFTVAGTNGELKNVALVOPQINGATASFTQVQAGYGTFRFTIGAGMG-- 235
Qy 239 AAKITRSFOAVGDIAPAYAVTVLIEIDTLRSSPAENKAKKRASLVGVSSTTFEYILC 298
Db 236 -----SFISIALFPFADGFRVTAAGLOSEMR-----EPKTPMALFLGLITTFEYIL 284
Qy 299 G-----CIGYAAFNNAPGDLTDFGFEPFPLIDPANACIAVHLIGAYQ 343
Db 285 AVALSINGLFGGMEBSWG-KLENNKRAQIV-FG-----VVMIMIGIVGLIIN 332
Qy 344 VRAQPIFOFVER-KCKRNPDKKFTITSESVVVPFLGKNISLFLVWRTAVVITVVA 402
Db 333 GPALMAPREVEDLLAQGDLPFKQVQGRINPKRPVVG-----VIVCLVLSLV 380
Qy 403 MTFPFPNALIGLIGASFPPLTVYFPVEHIAQTIKK-----YS-----ARMIATKT 450
Db 381 QV-----LFTVIGALAVIPTVADY---KNYVTEIDKLSNQMLYSPSDLMATWTSLE 431
Qy 451 MCVYCLIVSLAAGSIAGLISSVKTYK 478
Db 432 FAFI-----ACAFGAIIVNRKTKK 450

RESULT 10
YEH4 YEAST STANDARD; PRT; 480 AA.
AC P39981;
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE Hypothetical 53.3 kDa protein in HXT8-CAN1 intergenic region.
GN YEL064C.
OS Saccharomyces cerevisiae (Baker's Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_Taxid=4932;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Dierrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA Aviles E., Betno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA Chung E., Duncan M., Guzman E., Hartzell G., Hunnicke-Smith S.,
RA Hyman R., Kayser A., Komp C., Laszkari D., Lew H., Lin D.,
RA Mosedale D., Nakahara K., Namach A., Norgren R., Oefner P., Oh C.,
RA Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA Taylor P., Wei Y., Yelton M., Botstein D., Davis R.W.;
RL Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: SOME, TO C.ELBEGANS F13H10.3.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

```
CC -----
DR EMBL; U18795; AAB65023.1; -
DR SGD; S0000790; YEL064C.
DR InterPro; IPR002422; AA/rel_prmease2.
DR Pfam; PF01490; Aa trans; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 72 92 POTENTIAL.
FT TRANSMEM 95 115 POTENTIAL.
FT TRANSMEM 145 165 POTENTIAL.
FT TRANSMEM 214 234 POTENTIAL.
FT TRANSMEM 263 283 POTENTIAL.
FT TRANSMEM 297 317 POTENTIAL.
FT TRANSMEM 338 358 POTENTIAL.
FT TRANSMEM 394 414 POTENTIAL.
FT TRANSMEM 447 467 POTENTIAL.
SQ SEQUENCE 480 AA; 53323 MW; 45799CDB3695B78E CRC64;

Query Match 5.0%; Score 125.5; DB 1; Length 480;
Best Local Similarity 20.1%; Pred. No. 0.078;
Matches 90; Conservative 75; Mismatches 214; Indels 69; Gaps 16;

QY 44 ASAHITAVIGSGVLSLAWAIAQLGWIAGTSILLIFSFTYFTSTMLADCYRAPDPVTGK 103
DB 72 AFWNLANSILGAGIITQPAFKNAGILGILLSVALGFIWDTLRLV----INLTLAGK 127

QY 104 RNYTMDVRSYLGGRKVQLC-----GVAQYGNLIGVTYGTITASISLVAVGKSNCFHDK 159
DB 128 R--TYGCTVEHVWGKGLLILPTNGLFAFGGCI---GVCIIIGDTIPHLRAIFSQND 181

QY 160 GHTADCTISNYPMAVFGIIQVILSQIPNFKLSFLSIAAAVMSFYATIGLAIATVA 219
DB 182 GNVHFLRNRVIVVMVTTTISPFLSKRNEALSFKASFLAVISM-----IIIVLTVVI 234

QY 220 GGVKVGKTSMTGTAVGV-DVTAQAKIWRSFQAVGDIAPAYAYATVLEIQDTLRRSSPAENK 278
DB 235 RGPMPLPYDNKSHLSKLSDFEMKATIFRSLSVI-SPALVCHNHTSFFPMSNRVSV----A 289

QY 279 AKKRASLVGVSTTFFYIILGCGICGYAAGNNAFGDFTDFGFPFWLIDFANACIAVHL 338
DB 290 KFTRLTHISIIISVICCALMGVSGFAVEKTKGNVLSNFPFGDGT--AINIARLCFGFNM 347

QY 339 IGAYQVFAQPIQFQVEK-----KCN--RNPYDNKFTISEYSVNVPLGFKNISLFLRL 388
DB 348 LITTF-----PMEIFVLVRDVGNNLLHCELNKYNDEHTQLSGQHV----- 387

QY 389 VWRVAVVLTIVVAMIFPPFNATLGLIGLGAASFVPLTVYPPVEMHIAQTIKKYARWIAL 448
DB 388 VITSSLVFTIMGLSLTLCNLGALFELIGATTASTMAYILPPYTNLLLTSSKKK---SW-KE 443

QY 449 KTMCVYCLIVSLAAAGSIAGLISSVKT 476
DB 444 RLPFFYLICI-----CFGFMIMISSQT 465

RESULT 11
FUII YEAST
ID FUII YEAST STANDARD; PRT; 639 AA.
AC P38196;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Uridine permease.
GN FUII OR YBL042C OR YBL0406.
OS Saccharomyces cerevisiae (baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S288c;
RX MEDLINE=955176707; PubMed=7871888;
```

```
RA de Wergifosse P., Jacques B., Jonniaux J.-L., Purnelle B., Skala J.,
RA Goffeau A.;
RT "The sequence of a 22.4 kb DNA fragment from the left arm of yeast
RT chromosome II reveals homologues to bacterial proline synthetase and
RT murine alpha-adaptin, as well as a new permease and a DNA-binding
RT protein."
RL Yeast 10:1489-1496 (1994).
RN [2]
RN CHARACTERIZATION.
RP MEDLINE=98146544; PubMed=9485596;
RX Wagner R., de Montigny J., de Wergifosse P., Souciet J.-L., Potier S.;
RT "The ORF YBL042 of Saccharomyces cerevisiae encodes a uridine
RT permease."
RL FEMS Microbiol. Lett. 159:69-75 (1998).
RN [3]
RN CHARACTERIZATION.
RP MEDLINE=20408929; PubMed=10827169;
RX Vickers M.F., Yao S.Y., Baldwin S.A., Young J.D., Cass C.E.;
RT "Nucleoside transporter proteins of Saccharomyces cerevisiae.
RT Demonstration of a transporter (FUII) with high uridine selectivity
RT in plasma membranes and a transporter (FUN26) with broad nucleoside
RT selectivity in intracellular membranes."
RL J. Biol. Chem. 275:25931-25938 (2000).
CC -!- FUNCTION: HIGH-AFFINITY TRANSPORT OF URIDINE.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein.
CC -!- SIMILARITY: BELONGS TO THE ALLANTOIN PERMEASE FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
```

```
EMBL; X78214; CA55059.1; -.
DR EMBL; Z35803; CA84862.1; -.
DR PIR; S45776; S45776.
DR SGD; S0000138; FUII.
DR InterPro; IPR001248; Cyt_pur_permease.
DR Pfam; PF02133; Transp_cyt_pur; 1.
DR TIGRFAMs; TIGR00800; ncs1; 1.
KW Transmembrane; Transport.
FT TRANSMEM 163 180 POTENTIAL.
FT TRANSMEM 201 225 POTENTIAL.
FT TRANSMEM 260 276 POTENTIAL.
FT TRANSMEM 284 305 POTENTIAL.
FT TRANSMEM 368 392 POTENTIAL.
FT TRANSMEM 417 435 POTENTIAL.
FT TRANSMEM 461 477 POTENTIAL.
FT TRANSMEM 484 507 POTENTIAL.
FT TRANSMEM 538 562 POTENTIAL.
FT TRANSMEM 573 590 POTENTIAL.
SQ SEQUENCE 639 AA; 72165 MW; 757325DEP4100399 CRC64;

Query Match 4.8%; Score 121.5; DB 1; Length 639;
Best Local Similarity 18.3%; Pred. No. 0.2;
Matches 101; Conservative 80; Mismatches 175; Indels 195; Gaps 28;

QY 35 EKRTGTWLTASAHITAVIGSGVLSLAWAIA-----QLG---WIAGTSILLIFSFTYF-- 85
DB 123 KRRTWTW---KQYIFFWISGSGFNVN-TWQISATGLQLGLNWNQWICWVGTTFVAFPLI 178

QY 86 TSTMLADCYRAPDPVTGKRN-YTYMDV-----VRSYLGGKRVQLCGVAQY 129
DB 179 LGSGKGNVNHISFPPISSRVSGFIYSIWIIVNRVVMACVMNSTLAYIGSQCVQLMLKAIF 238

QY 130 GNLIGVTYGTITASISLVAVGKSNCFHDKHTADCTISNYPMA--VFGIIQVILSQIP 187
DB 239 GTNLNTRIKDTI-----KNPNLTNFEFCFMVFWVACLPLFWFP 277

QY 188 NFKLSFLSIAAAVMSFYATIGLAIATVAGGK-----VGKTSMTGTAVGVVDVTAQKI 243
```

```

Db 278 PDKRHFPLKSAITPFA-----AFGLIWLTLCKAKGHALGSLINDNGAISKTVLA----- 329
Qy 244 WSPQAVGDIAPAYAVAYVLIIEIOTLRSSPAENKAMKR-----ASLVGVTSTT 292
Db 330 MSVIRAL--MSALDNSTLILNAPDTRFGKTKSSVQGLALPVCYALISLIGLSTVS 387
Qy 229 FFIYILGCGTIGY-----AAFGNAPGDDPLTDGPF-PEEP----- 324
Db 388 AAYTYLGG--VNYMSPDLILNRYLDNYTSGNRA--GVFLISFPAFDQGANLISGNSIPAGTD 445
Qy 325 ---WLIDPAN-----ACIAVHL-----IGAYVFAQPIF-----QF 352
Db 446 LTRALLPKFPIINIRGSYICALLISLAIPTPMDLSSSSFTTALAAVAVLSAIGVISADYF 505
Qy 353 VEKK-----CNRNYPDNKKFTITSEYVAVPPLGKFNISLFLVWRTAVYVITTVAMTF 405
Db 506 IYRKGVVNI-FHCCTDKRSGSYVMNKKGTN-----WR-----AVAYITF 543
Qy 406 ---PPFNAILGLIG-----AASFVPLTVYFPEVMHIAQTKI-- 438
Db 544 GIAPNFAGFLGSGVSVPIGAMKVVYLYNFVGYLLAALSYCLIVFYPIKGIIPGDAKTD 603
Qy 439 KKYSANWMLK 449
Db 604 RKMLEEVEVE 614

RESULT 12
YFYA_SCHPO STANDARD; PRT; 591 AA.
ID YFYA_SCHPO
AC Q9UT18;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative amino-acid permease C9.10.
GN SPAC9.10.
OS Schizosaccharomyces pombe (fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
NCBI_TaxID=4896;
RX MEDLINE=21848401; PubMed=11859360;
RC STRAIN=972;
RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA Sgourou J., Peat N., Hayles J., Baker S., Baeham D., Bowman S.,
RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA Mooney F., Mould S., Mungall K., Murphy L., Niblett D., Odell C.,
RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,
RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA Woodward J., Volckaert G., Aert R., Robben J., Grymptez B.,
RA Weljens J., Vanstreels B., Rieger W., Schaefer M., Mueller-Auer S.,
RA Gaboli C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA Borym K., Langer I., Beck A., Lehnach H., Reinhardt R., Pohl T.M.,
RA Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelure V., Mottier S.,
RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,
RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.,
RT The genome sequence of Schizosaccharomyces pombe.
RL Nature 415:871-880(2002).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AL121764; CAB57428.1; -
DR InterPro; IPR002293; AA/re1_prmaseel.
DR InterPro; IPR004840; AA/permease.
DR InterPro; IPR004841; permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; PS00218; AMINO ACID PERMEASE 1; FALSE NEG.
KW Hypothetical protein; Transport; Amino-acid transport;
KW Transmembrane
FT TRANSMEM 98 118 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 397 417 POTENTIAL.
FT TRANSMEM 450 470 POTENTIAL.
FT TRANSMEM 545 565 POTENTIAL.
SQ SEQUENCE 591 AA; 65069 MW; 7FCCDD707AC7849D4 CRC64;

Query Match 4.8%; Score 121; DB 1; Length 591;
Best Local Similarity 22.0%; Pred. No. 0.2;
Matches 108; Conservative 70; Mismatches 194; Indels 120; Gaps 27;

Qy 55 SCV-LSLAWAIQOLGNIAGTSLILIFSTYTSMLADCYAPDPVGRKNTYMDVR 113
Db 94 SGVRLTFSGISFGPAAWMSAMLVGFCISYTAACLAEICSL-PAAGS---IYLMAE 149
Qy 114 SYLG---GRKVLGCAVAGNIGVTVGTTTASISLVAVGRSGNCPHDGHTADCTISY 170
Db 150 S-AGPRGRFVSFL-VAMWSTTAMTTFVASTITOSTANFPAEAVSTFNNWPTND---SDV 204
Qy 171 PYMAVFGII-----QVILSQI-PNFHLSFLSMAAVMSFTYATIGLAIAATVA--- 219
Db 205 KPRAVQWIVAELVLVFTLLNQVPPRYWKIFKAW-ILMFIDYWNIIWVPVATSKRD 263
Qy 220 GGRVGTSMTG-----AVGVDTAAQCI---WR---SFOAVGDIAPAY-A 258
Db 264 GFRSAKWVETETIYDQAGYKEVDANGNPILASLSIVPKGQWCLSYFATGVIYGYDA 323
Qy 259 YATVLIETIDTLRSSPAENKAMKRASLVGSTTFP-----YILGCGTIGYAAF--- 306
Db 324 SGHIAETKDA-----STKAARGLFYSTVTSIYAFSLAILVLC-CPDLDTFTAI 373
Qy 307 ---GNNAFGDLTDGFPF-----EPFWLIDPANACIAVHLIGAYVFA-----OP 348
Db 374 LYNDSPOFVAVFYSLGRGHVAVNVIIIEIFLNGVSV-LACSRVFPVSRGVLP 432
Qy 349 IFQVPEKKCNRNYPDNKFTITSEYVAVPPLGKFNISLFLVWRTAVYVITTVAMTFP-- 406
Db 433 FSNWISQVSKTQCPKXAI-----TVIYISALDLCTLLPBA 468
Qy 407 -FENAILGIGAASFVPLTVYFPEVMHIAQTKIKKYSARWIALKTMCYCLIVSLAAG 465
Db 466 VAFSTISAAGAPFAAVAVLAFCRILFTRDKPK--GRW-SLGWLSKPLVITTVY--- 522
Qy 466 SIAGLISVKTQ 477
Db 523 NLFALVNVSPY 534

RESULT 13
YV39_MYCTU STANDARD; PRT; 440 AA.
ID YV39_MYCTU
AC Q10858;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical transport protein Rvi1999c.

```

GN RV1999C OR MT2055 OR MTCV39.19.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN (1) SEQUENCE FROM N.A.
RP STRAIN=H37Rv;
RC MEDLINE=98295987; PubMed=9634230;
RX Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekai F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sultun J.B., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).
RN (2) SEQUENCE FROM N.A.
RP STRAIN=CDC 1551 / Oshkosh;
RC Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Uitterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bishai W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains.";
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: PROBABLE AMINO-ACID OR METABOLITE TRANSPORT PROTEIN.
CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -!- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; Z74025; CA998388.1; --
DR EMBL; AE007057; AAK46332.1; --
DR TIGR; MT2055; --
DR Tuberculist; RV1999C; --
DR InterPro; IPR002293; AA/rel_permease1.
DR InterPro; IPR002422; AA/rel_permease2.
DR InterPro; IPR004840; AAC_permease.
DR InterPro; IPR004841; Permease.
DR Pfam; PF00324; aa_permeases; 1.
DR PROSITE; P500218; AMINO ACID PERMEASE 1; FALSE NEG.
KW Hypothetical protein; Transport; Transmembrane; Complete proteome.
FT TRANSMEM 24 44 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 93 113 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
FT TRANSMEM 183 203 POTENTIAL.
FT TRANSMEM 229 249 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 346 366 POTENTIAL.
FT TRANSMEM 379 399 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 440 AA; 45267 MW; 9592EE0A842D18CD CRC64;
SQ SEQUENCE 440 AA; 45267 MW; 9592EE0A842D18CD CRC64;
Query Match 4.8%; Score 120; DB 1; Length 440;
Best Local Similarity 21.9%; Pred. No. 0.18;
Matches 95; Conservative 52; Mismatches 176; Indels 110; Gaps 20;
QY 24 DPTKNVDEGKREKTCWTLTASAHITAVIGSGVLSLAWAIAQLGWIAGTSILL----- 77

Db 6 DDPDPE--LRRRLGL-LDAVWIGLSMIGAFPA---ALAPRAYAGSGLLGLAVAA 59
QY 78 IPSFTYTTSTMLADCYRAPDPVTKRNYTMDVRSYLGGRKVQLCGVAQVGNLIGVTV 137
Db 60 VVAYCNAISSARLAARY---PASG-GTYVY-----GRMRLGDFWGYLA 98
QY 138 GYTTTASISLVAVGKSNCFHDKGHTADC-----TISNYPY-----MAVFGIIQVILSQI 186
Db 99 GMGFVV-----GKTASCAAMALTGVFVWPAQAHAAVAVVAVVAVTAVNY 142
QY 187 PNFHKLSEL--SIMAAVMSFTYATIGIGLAITATVAGGKVGKTSMTGTAVGVVDVTAQKIW 244
Db 143 AGIQKSAWLTSIVAVL-----VLTAVVAAYAGSGAADPARLDIGVDA-----HWV 190
QY 245 RSFQAVGDIAPAYA-----YATVLIBIQOTLRSSPAENKAMKRAVLGVSTTTFFFYILCGC 300
Db 191 GMLQAAGLFFFAFAGYARIATLGEVVRDPARTIP-----RAIPALGITLAVVALVAV 243
QY 301 IGYAAFGNN-----AP-GDFLTDFGFFEPFWMLIDFANACIAVHLIGAVQVFAQPIFOFV 353
Db 244 AVIIVLGPQRLARAAPLSEAMRVAGV---NWLIPVVOIGAAVAAIGSLALLILGVSRIT 300
QY 354 EKKCKNRNTPDNKFKITSEYSVNVVPLGKFNISLPLRLVMTYAVVITTVVAMIFPPFFNAILG 413
Db 301 LAMARDRLHPLRWLAHAVHPRFKVPF-----RAELVVGAVVAALAAATAD--IRG 345
QY 414 LIGAAASFVPLTVY 426
Db 346 AIGFSSFGVLVY 358
RESULT 14
YWCA_BACSU STANDARD; PRT; 516 AA.
ID YWCA_BACSU STANDARD; PRT; 516 AA.
AC P39599;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative symporter ywca.
OS YWCA OR IPA-31R.
GN Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN (1) SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98020537; PubMed=7934828;
RA Glaeser P., Kunst F., Arnaud M., Coudart M.P., Gonzales M.,
RA Hullo M.F., Tonescu M., Lubochinsky B., Marcelino L., Moszer I.,
RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN (2) SEQUENCE FROM N.A.
RP STRAIN=168;
RC MEDLINE=98044033; PubMed=9384377;
RA Kunst P., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell I.F., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enriot K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Coffeau A., Golightly E.J., Grandi G.,
RA Gutsepp G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,

RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogawara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Porroliot S., Prescott A.M.,
RA Prescen E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone P.,
RA Setiguchi J., Sekowska A., Seror S.D., Serro P., Shin B.S., Soldo B.,
RA Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni K.,
RA Tosoato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambolt R., Wandler E., Wiedler H., Weitzneger T.,
RA Witters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256 (1997).
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: BELONGS TO THE SODIUM: Solute symporter family (SSF).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73124; CA51587.1; -
DR EMBL: Z99123; CAB15850.1; -
DR Subtilist; Bg10577; ywca.
DR InterPro; IPR001734; Na/solut_symport.
DR Pfam; PF00474; SSF; 1.
DR TIGRfam; TIGR00813; ssa; 1.
DR PROSITE; PS00455; NA_SOLUT_SYMPT_1; 1.
DR PROSITE; PS00457; NA_SOLUT_SYMPT_2; 1.
DR PROSITE; PS00283; NA_SOLUT_SYMPT_3; 1.
KW Hypothetical protein; Transport; Transmembrane; Sodium transport;
KW Symport; Complete proteome.
FT TRANSMEM 7 27 POTENTIAL.
FT TRANSMEM 51 71 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 132 152 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 180 200 POTENTIAL.
FT TRANSMEM 236 256 POTENTIAL.
FT TRANSMEM 276 296 POTENTIAL.
FT TRANSMEM 323 343 POTENTIAL.
FT TRANSMEM 377 397 POTENTIAL.
FT TRANSMEM 398 418 POTENTIAL.
FT TRANSMEM 427 447 POTENTIAL.
FT TRANSMEM 465 485 POTENTIAL.
SQ SEQUENCE 516 AA; 54958 MW; E39628B26AA952D CRC64;
Query Match 4.7%; Score 118.5; DB 1; Length 516;
Best Local Similarity 19.3%; Pred. No. 0.26;
Matches 93; Conservative 73; Mismatches 216; Indels 100; Gaps 19;
QY 34 REKRGTWLTASAHITTAIVAGSGVSLMAIAQLGIAGTSLILFSTTYPTS-----T 88
DB 29 RKTSTSEFTAGGGLTGVNGLAIDVMSAASFLGIAG--MALYCGDFGFFSISFLVA 86
QY 69 WLADCVRAPDPTGKKNYTYMDVRSYLGGRKQVLCGVAGVQGLIGVTGYTTASISLV 148
DB 87 YLVVLYTVAEPLRNKGYTMADMTIARFEKRR--GVALL-NTIATSTPYMAQLVGVAG 143
QY 149 AVGKSNCFPHDGHDTADCTISNYPYMA---VFGLIYVILSQIPNFHLISFLSMAAVMSFT 205
DB 144 ALIK-----LLGSDYTAALVILGVLMITIVYFGGMIMTSVQIIKAVILMA 190
QY 206 VATTIGGLAIVAVGKVG--KTSMTGTAVGVDTVAQKIMRSFOAVGDIAFAYVAT- 261
DB 191 -GTLVISIIVFSKFGPSLNTMFQOMKTATPLGADFLNPGNKYVPLETSLNIALVLGTA 249

QY 262 ---VLIIQDTLRSSPAENKAKRASLVGVTTFEPILGCGIGY--AAF----- 306
DB 250 GLPHILIFFTVVKAKTARTSVASATWIIQV-----FYIMTVFLGFGAAAFGPDATTA 304
QY 307 ---GNNA-----GPELDFGFEFEPWLIIDPAN--ACIAVLHIGAVQVPAQIIPQFV 353
DB 305 DAGNMAAPLAKALGQFL--FAVVS--AIAFTIILAVTGLVLSAASAFADHYSQI 359
QY 354 EKKCRANYPDNKFTISEYVNVPLGKENISLFLWRTAVVITTVAMIF-----PF 407
DB 360 IRKGATEKE-----QMKARMAVSVAVSISILAIPOSLNAAF 399
QY 408 FNAILGLGAASFWPLTYVFPVENHIAQTKIKKYSARWIAIKTCVCYLVSLLAAAGSI 467
DB 400 LVALLAFVAASANLPLIVF-----TFPWKRFVAGSLGSLTGLISALVLSMSPSVM 454
QY 468 AG 469
DB 455 AG 456
RESULT 15
ID YSPK CAEEL STANDARD; PRT; 615 AA.
AC Q19425;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein F13H10.3 in chromosome IV.
GN F13H10.3.
OS *Caenorhabditis elegans*.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderidae; *Caenorhabditis*.
OX NCBI_TaxID=6229;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Cottage A.;
RL Submitted (JAN-1996) to the EMBL/Genbank/DBJ databases.
RN [2]
RP REVISIONS.
RA Durbin R.;
RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1- SIMILARITY: SOME TO YEAST YEL064C.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z68748; CA92933.2; -
DR WormPep; F13H10.3; CE27959.
DR InterPro; IPR002422; Aa/re1_pmease2.
DR Pfam; PF01490; Aa_trans; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 166 186 POTENTIAL.
FT TRANSMEM 195 215 POTENTIAL.
FT TRANSMEM 248 268 POTENTIAL.
FT TRANSMEM 342 362 POTENTIAL.
FT TRANSMEM 368 388 POTENTIAL.
FT TRANSMEM 416 436 POTENTIAL.
FT TRANSMEM 452 472 POTENTIAL.
FT TRANSMEM 496 516 POTENTIAL.
FT TRANSMEM 536 556 POTENTIAL.
FT TRANSMEM 562 582 POTENTIAL.
FT TRANSMEM 595 615 POTENTIAL.
SQ SEQUENCE 615 AA; 68692 MW; 5092F2ABF4A3A420 CRC64;

Query Match	4.6%	Score 115.5;	DB 1;	Length 615;
Best Local Similarity	18.5%;	Pred. No. 0.52;		
Matches	90;	Conservative	81;	Mismatches 164; Indels 151; Gaps 21;
Qy	48	IITAVIGSVLSAWATAQLGWIAGTSTLLIFSPITYFTSMMLADCYR----	APDPVTG	102
Db	172	IWNWMTGSLIAMPALWQAQGLVLIILMSNAALCFYTAIVIVIESPKRLQDLSVDPLLA	231	
Qy	103	KRNYYMDVRSYLGGRKVQLC-----GVAOYGNIJ-----	GTVGYGTITAS	144
Db	232	E----FSDVCKS-LFGRIGECAVVFSVCVLGGVIVYWLMNSNFLYYTGAAVYESMQPN	286	
Qy	145	ISLVAVGKSNCF-----	HDKGHTADCTISNPYMAVF	176
Db	287	STTIPVMENKTFTCDVYCPEQTQSOWTIPOWEKOLYDAVSEMEGETGDSSWDFKFWTLR	346	
Qy	177	GIIOVILLS---Q:PNFHKLSP-----LSTMAAAMSFYATIGIGLAIATVAGGKV	223	
Db	347	GTVPILYAFALPDLNFKSPFTTFKNVLGTYISWYLUM-FVF-----	388	
Qy	224	GKTSMTGTAVGDVDTAAQKIW-----RSFOAV-GDIAPAYAYATVLTIEIQTLRSSP-A	275	
Db	389	--SKLLECGVNMDPSNPKSIHYVOLANMHFPALSGTLTLSYFIHNAVITI---	LNRQKHP	443
Qy	276	ENKAMKASRLVGVSSTTTFFYLGCIGYAAP---GNNA PGDFLD FGFEPFWILDFA NA	332	
Db	444	ENNA--RDLSIGYCLVAFCCYVFIGFTFPAAPPVORSICISDNFLANFG-----	AGDVLS	495
Qy	333	CIAVHLIGAYQVPAQPIPOFEKKCNRPYPDNKFTITSEYSNVVFLGKFNLSLRLVWRT	392	
Db	496	TARLFLL-FQMITVLP LLMFLVR-----	SOLFVAFIGQTWPG	531
Qy	393	AyvittvAMI-----PPFNAILGLIGAASFwPLTWYPVPEVKHIAQ-----	T	436
Db	532	AIRVIILNVLIIAVAGPATPYPNVGSILRYVGSISGLVYVFPALPAMYVIKQSEAAGTLT	591	
Qy	437	KIKKYS	442	
Db	592	PMKKYA	597	

Search completed: November 22, 2002, 13:30:18
Job time : 17 secs